

Departamento de Biología Molecular  
Facultad de Ciencias  
Universidad Autónoma de Madrid

Especificidad en la interacción de la proteína p4 con el DNA y formación del complejo regulador de la transcripción del bacteriófago  $\phi 29$ .

**TESIS DOCTORAL**

*Laura Pérez García*

Madrid, 2007

Departamento de Biología Molecular  
Facultad de Ciencias  
Universidad Autónoma de Madrid

Especificidad en la interacción de la proteína p4 con el DNA y formación del complejo regulador de la transcripción del bacteriófago ø29.

Memoria presentada por la Lda. en Farmacia y Bioquímica *Laura Pérez García* para optar al grado de Doctora en Ciencias por la Universidad Autónoma de Madrid.

*Madrid, Febrero 2007*

El trabajo presentado en esta Tesis Doctoral ha sido realizado en el Centro de Biología Molecular “Severo Ochoa” (CSIC-UAM) bajo la dirección de la Dra. Ana Camacho Pedrero. Para la realización de este trabajo se contó con una beca de Formación de Personal Investigador de la Comunidad Autónoma de Madrid.

# ÍNDICE

ABREVIATURAS .....	3
SUMMARY .....	4
INTRODUCCIÓN.....	5
1.- TRANSCRIPCIÓN EN ORGANISMOS PROCARIOTAS.....	5
1.1) <i>RNA polimerasa de bacterias</i> .....	5
1.2) <i>Promotores</i> .....	9
1.3) <i>El proceso de transcripción. La iniciación.</i> .....	10
1.4) <i>Modulación de la iniciación de la transcripción</i> .....	12
- 1.4.1 <i>Secuencias de los promotores</i> .....	12
- 1.4.2 <i>Regulación por factores <math>\sigma</math> alternativos</i> .....	12
- 1.4.3 <i>Pequeños ligandos</i> .....	13
- 1.4.4 <i>Factores de transcripción</i> .....	13
1.4.4 a) <i>Mecanismos Simples de Regulación mediados por factores de transcripción</i> .....	13
1.4.4 b) <i>Mecanismos Complejos de Regulación mediados por factores de transcripción</i> .....	15
2.- TRANSCRIPCIÓN EN EL BACTERIOFAGO $\phi 29$ .....	18
2.1) <i>Mapa genético y transcripcional de <math>\phi 29</math>.</i> .....	18
2.2) <i>Proteína p4</i> .....	21
- 2.2.1 <i>Región de los promotores centrales y sitios de unión de p4</i> .....	21
- 2.2.2 <i>Mecanismo de regulación <i>in vitro</i> mediado por la proteína p4</i> .....	22
2.3) <i>Proteína p6</i> .....	23
2.4) <i>Complejo nucleoproteico p4-p6</i> .....	23
OBJETIVOS.....	25
MATERIALES Y MÉTODOS.....	26
1.- MATERIALES.....	26
2.- MÉTODOS .....	29
2.1) <i>Clonajes</i> .....	29
2.2) <i>Mutagénesis dirigida</i> .....	30
2.3) <i>Purificación de proteínas</i> .....	31
2.4) <i>Existencia de curvatura intrínseca en el DNA</i> .....	32
2.5) <i>Permutación circular</i> .....	32
2.6) <i>Ensayos de modificación de la movilidad electroforética del DNA</i> .....	34
2.7) <i>Ensayos de protección a la digestión del DNA por DNasa I.</i> .....	34
2.8) <i>Secuenciación del DNA</i> .....	35
2.9) <i>Ensayos de transcripción truncada "in vitro"</i> .....	35
RESULTADOS.....	36
1.- DETERMINACIÓN DEL SITIO MÍNIMO DE UNIÓN DE P4 .....	36
2.- AFINIDAD RELATIVA DE P4 POR SUS SITIOS DE UNIÓN .....	38
3.-CONFORMACIÓN DEL DNA DE LA REGIÓN 1 .....	39
3.1) <i>Curvatura intrínseca</i> .....	39
3.2) <i>Curvatura inducida</i> .....	40



3.3) Influencia de la secuencia de cada sitio en la inducción de la curvatura por p4 en la región 1 .....	42
3.4) Curvatura inducida por el complejo p4-p6 .....	44
4.- MUTANTES DE P4 DE UNIÓN AL DNA .....	46
4.1) Mutantes de p4.....	49
4.2) Unión al DNA .....	49
4.3) Regulación de la transcripción.....	50
4.4) Formación de complejos con las proteínas p4 mutantes .....	51
- 4.4.1 Complejos p4-RNAP-DNA .....	51
- 4.4.2 Complejos p4-p6-DNA.....	53
4.5) Capacidad de inducir una curvatura en el DNA .....	55
5.- ESTUDIO DE LOS DETERMINANTES DE ESPECIFICIDAD DE LA SECUENCIA DE UN SITIO DE UNIÓN DE P4 .....	58
5.1) Modificaciones en el sitio 1 .....	59
5.2) Modificaciones en el sitio 2 .....	60
5.3) Modificaciones en el sitio 4 .....	60
5.4) Modificaciones en el sitio 3 .....	62
- 5.4.1 Modificaciones en el tamaño del sitio .....	62
- 5.4.2 Modificaciones en la secuencia del sitio 3 entre las posiciones 8-12 ...	63
- 5.4.3 Diferencias en los extremos del sitio 3 .....	65
6.- ESTUDIO DE LA IMPORTANCIA DEL NÚMERO DE SITIOS Y DE LA SECUENCIA EN LOS EXTREMOS DE CADA SITIO. ....	67
6.1) Sitio 1.....	69
6.2) Sitio 2.....	70
6.3) Sitio 3.....	72
6.4) Sitio 4.....	73
7.- COOPERATIVIDAD EN LA UNIÓN DE P4 A SUS SITIOS .....	75
8.- ESTUDIO DE LA REGULACIÓN DE LA TRANSCRIPCIÓN DEL FAGO Nf RELACIONADO CON Ø29 .....	77
8.1) Interacción de las proteínas p4 de Nf y p6 de Nf en la región de los promotores A2c, A2b y A3 .....	78
8.2) Determinación de los sitios de unión de p4 de Nf en su genoma. ....	80
8.3) Análisis de la competición de la RNAP con los complejos p4-p6, homólogo y heterólogo .....	81
8.4) Regulación de la transcripción "in vitro" del DNA de Nf.....	82
<b>DISCUSIÓN .....</b>	<b>85</b>
1.- CONFORMACIÓN DEL DNA EN LA SECUENCIA DE LOS PROMOTORES A2c, A2b Y A3.....	85
2.- INTERACCIÓN DE P4 CON SUS SITIOS DE UNIÓN .....	86
3.- FUNCIÓN DE LOS SITIOS DE UNIÓN DE P4 EN EL DNA.....	95
4.- GENERALIDAD Y DIVERSIDAD DEL MECANISMO DE REGULACIÓN DE LA TRANSCRIPCIÓN.....	99
<b>CONCLUSIONES .....</b>	<b>100</b>
<b>BIBLIOGRAFÍA.....</b>	<b>101</b>

# *ABREVIATURAS*

A: adenina	LB: medio de cultivo Luria Bertani
aa: aminoácido	mRNA: ácido ribonucleico mensajero (del inglés, siglas <i>messenger ribonucleic acid</i> )
A/T: adenina(s)/timina(s)	NTD: dominio amino terminal (del inglés, siglas <i>N-terminal domain</i> )
ATP: adenosina- 5'-trifosfato	nt: nucleótido
BSA: seroalbúmina bovina (del inglés, siglas <i>bovine serum albumin</i> )	NTP: nucleosido-5'-trifosfato
βME: β- mercaptoetanol	OD: densidad óptica (de inglés, siglas <i>optical density</i> )
C: citosina	PCR: reacción en cadena de la polimerasa (del inglés, siglas <i>polymerase chain reaction</i> )
oc: complejo abierto (del inglés, siglas <i>open complex</i> )	PEI: polietilenimina
cc: complejo cerrado (del inglés, siglas <i>close complex</i> )	pb: par(es) de bases
CRP: proteína receptora de cAMP (del inglés, siglas <i>cAMP receptor protein</i> )	poli[d(I-C)]: polímero de deoxiinosina y deoxicitidina
CTD: dominio carboxilo terminal (del inglés, siglas <i>C-terminal domain</i> )	RNA: ácido ribonucleico (del inglés, siglas <i>ribonucleic acid</i> )
α-CTD: extremo carboxilo terminal de la subunidad α de la RNAP	RNAsin: inhibidor de RNAsas de tipo A
CTP: citidina-5'-trifosfato	RNAP: RNA polimerasa (del inglés, siglas <i>ribonucleic acid polymerase</i> )
dATP: 2'-desoxiadenosina- 5'-trifosfato	rRNA: ácido ribonucleico ribosómico (del inglés, siglas <i>ribosomal ribonucleic acid</i> )
DAP: 2- diaminopurina	PAGE: electroforesis en gel de poli(acrilamida) (del inglés, siglas <i>polyacrilamide gel electrophoresis</i> )
DNA: ácido desoxirribonucleico (del inglés, siglas <i>deoxyribonucleic acid</i> )	SDS: dodecil sulfato sódico (del inglés, siglas <i>sodium dodecyl sulphate</i> )
DNasa I: desoxirribonucleasa I	tRNA: ácido ribonucleico de transferencia (del inglés, siglas <i>transfer ribonucleic acid</i> )
dNTPs: 2'-desoxirribonucleósidos- 5'-trifosfato	T: timina
DTT: ditioneitol	Tris: trihidroximetil aminometano
EDTA: ácido etilen diamino tetraacético	UDG: uracil DNA glicosilasa
Fis: factor estimulador de la inversión (del inglés, siglas <i>factor for inversion stimulation</i> )	UTP: uridina- 5'- trifosfato
G: guanina	wt: tipo salvaje o silvestre (del inglés siglas <i>wild type</i> )
GTP : guanosina-5'-trifosfato	
H-NS: proteína tipo histona de estructuración del nucleóide (del inglés, siglas <i>histone-like nucleoid structuring</i> )	
HU: proteína nucleóidea inestable por calor (del inglés, siglas <i>heat-unestable</i> )	
IHF: factor de integración del hospedador (del inglés, siglas <i>integration host factor</i> )	
IPTG: isopropil-β-D-tiogalactósido	

# *SUMMARY*

Gene expression of bacteriophage  $\phi 29$  is carried out by the *Bacillus subtilis* RNA polymerase and is regulated by early viral proteins. The switch from early to late transcription involves the formation of an extended nucleoprotein complex of proteins p4 and p6 in the 219 bp sequence of DNA comprising A2c, A2b and A3 promoters. In this work we analyzed the determinants of the specificity of p4-DNA interaction, the role of the p4 binding sites and the regulation of transcription in a  $\phi 29$ -related phage.

Early transcription produces proteins implicated mainly in the synthesis of  $\phi 29$  DNA and transcription regulators; late transcription gives rise to structural proteins required for production of the infective viral particles.

Protein p6 is a nucleoid protein, small, dimeric, and very abundant in infected cells; nucleoid proteins bind DNA non-specifically, although some of them show a preference for certain sequences or structures of DNA.

Protein p4 is the phage transcriptional factor. It binds to the DNA at two regions. Region 1 is located between early promoters A2c and A2b and region 2 is located between promoter A2b and the late promoter A3. Each region has two binding sites; p4 binds as a dimer to each binding site.

By means of directed mutagenesis of different residues of the protein and by modification of nucleotides at the binding sites we could define the principal characteristics of a p4 binding site and the determinants of the specificity on the p4-DNA complex; we could observe, moreover, that p4 does not present the same affinity either for its binding sites or for the sequence at each end of the binding site. These results lead us to propose a zipper mechanism for the binding of p4.

Here we demonstrate that p4 does not require an intrinsically bent DNA for binding and that whenever p4 binds to a complete region, it bends the DNA and induces a curvature around 86°.

Studies of the role of each binding site allowed us to conclude that sites 2 and 4 do not have a direct implication in the regulation of the switch from early to late transcription. The functional complex is formed by binding of p4 at sites 1 and 3 and the sequence comprised between them is covered by protein p6, being this complex able to repress early transcription and simultaneously activate the late transcription. This mechanism is conserved in another *B. subtilis* phage, Nf, a phage from a different group of the  $\phi 29$  family. Nf bacteriophage encodes proteins p4N and p6N, homologous to  $\phi 29$  early proteins p4 and p6, respectively, and contains two early promoters homologous to A2c and A2b and a late promoter similar to A3. Transcriptional switch also requires the formation of a p4-p6 nucleoprotein complex. However,  $\phi 29$  proteins are unable to regulate the Nf promoters, suggesting that the transcriptional regulation mechanism of each phage has evolved differently.

# *INTRODUCCIÓN*

Un organismo vivo se caracteriza por su capacidad para reproducirse y transmitir la información genética a su descendencia con el fin de perpetuar la especie. Los procesos biológicos que, fundamentalmente, permiten que esto ocurra se pueden reducir a tres: *la replicación, la transcripción y la traducción*.

*La replicación* es el proceso por el cual se duplica la información genética de un organismo. En la mayoría de los casos, el DNA es la molécula que contiene el material hereditario y una copia completa del mismo debe pasar de una generación a la siguiente.

*La transcripción* es el proceso mediante el cual se copia una cadena del DNA de doble banda para producir una cadena sencilla de RNA. La molécula de RNA es idéntica en secuencia a una de las bandas del DNA (la codificante) y complementaria a la otra (la molde). Los productos de la transcripción pueden actuar directamente como moléculas estructurales o funcionales (rRNAs o ribozimas), o bien pueden ser moléculas transitorias portadoras de información para la síntesis de proteínas (mRNAs).

*La traducción* es el proceso a través del cual la información contenida en una molécula de mRNA se convierte en una secuencia de aminoácidos que, tras plegarse en una determinada estructura tridimensional, da lugar a una proteína.

## **1.-TRANSCRIPCIÓN EN ORGANISMOS PROCARIOTAS**

Los organismos procariotas se consideran los más simples y primitivos. Las bacterias poseen más de 4000 genes, siendo necesario que los procesos de replicación, transcripción y traducción estén sincronizados de tal modo que, en un determinado momento, se expresen unos genes en concreto a los niveles adecuados. Tiene que existir un control muy fino, cualitativo y cuantitativo, de la síntesis de macromoléculas en la célula. Los organismos procariotas están sujetos a múltiples cambios en el entorno que los rodea y para adaptarse a las nuevas situaciones, deben producir diferentes grupos de moléculas empleando para ello sistemas de regulación, en ocasiones, muy complejos. El paso central en la regulación de la expresión de genes dentro de una cascada de reacciones es **la transcripción**.

La reacción de transcripción la lleva a cabo una enzima llamada *RNA polimerasa dependiente de DNA* (RNAP) y consiste en la formación de enlaces fosfodiéster entre sustratos de ribonucleósidos trifosfato, siguiendo la secuencia de una cadena molde de DNA. Las secuencias que señalizan en el DNA dónde debe tener lugar el inicio de la transcripción se llaman promotores.

El *proceso de transcripción* se puede dividir en tres etapas: iniciación, elongación y terminación. De forma simplificada, el proceso comienza con la interacción de la RNAP con el promotor, dando lugar al complejo cerrado de transcripción (cc). A continuación, tiene lugar la separación de las dos cadenas del DNA, mediada por cambios conformacionales de la RNAP, formando el complejo abierto de transcripción (oc). Tras el escape de la enzima del promotor y la elongación de la cadena del RNA naciente, se forma el complejo ternario de elongación que, cuando encuentra secuencias terminadoras en el DNA, se disocia del mismo, alcanzándose el final del proceso de transcripción o terminación.

### **1.1) RNA polimerasa de bacterias**

La RNAP es la enzima central en la transcripción y expresión de genes y es, además, la diana más frecuente para la regulación del proceso. La forma multimérica más simple de RNAP es la de bacterias, que consiste en un enzima mínimo o *core* de aproximadamente 400 KDa, compuesto por cinco polipéptidos ( $\alpha_2\beta\beta'\omega$ ) (Burgess,

1969), que se asocia con un factor  $\sigma$  para formar la *holoenzima*. El *core*, contiene toda la maquinaria catalítica requerida para la síntesis de RNA; pero sólo la *holoenzima* puede iniciar la transcripción en sitios específicos, ya que el *core* por sí mismo es incapaz de reconocer los promotores. Es, por tanto, la subunidad  $\sigma$  la responsable del reconocimiento de la secuencia que indica desde donde va a tener lugar el proceso de transcripción (Burgess *et al.*, 1969).

La secuencia, estructura y función del *core* están conservadas evolutivamente desde las bacterias al hombre (Allison *et al.*, 1985; Borukhov y Nudler, 2003; Ebright, 2000; Minakhin *et al.*, 2001; Sweetser *et al.*, 1987).

La RNAP de *Escherichia coli* es la mejor caracterizada tanto genética como bioquímicamente, pero su estructura sólo se ha conseguido, mediante microscopía crioelectrónica, a una relativa baja resolución (Darst *et al.*, 2002). Actualmente se dispone de las estructuras a resolución atómica del *core* de la RNAP de *Thermus aquaticus* (*Taq*) a 3,3 Å (Zhang *et al.*, 1999) y de la *holoenzima* de *Thermus thermophilus* (*Tht*) a 2,6 Å (Vassylyev *et al.*, 2002). Existe homología entre las secuencias de las subunidades de las RNAP de *Taq* y *Tht* con las de *E. coli* y *Bacillus subtilis*, así como una amplia coincidencia en sus funciones, de modo que los datos estructurales obtenidos para las primeras, pueden ser aplicados de forma general a las segundas (Borukhov y Lee, 2005; Darst, 2001; Gruber y Gross, 2003). En los últimos años, ha tenido lugar un avance espectacular en los estudios estructurales de la RNAP, gracias a la resolución de la estructura de la enzima formando complejo con el DNA (Murakami *et al.*, 2002), de las estructuras de diferentes dominios de las subunidades de la RNAP y de las subunidades formando complejos con factores de transcripción y con el DNA (Benoff *et al.*, 2002; Campbell *et al.*, 2002; Zhang y Darst, 1998).

La estructura de las RNAP recuerda a la pinza de un cangrejo (figura 1); las dos subunidades mayores del *core*,  $\beta$  y  $\beta'$ , forman los brazos superior e inferior de la pinza, respectivamente. Estos brazos se encuentran unidos por detrás a través de los dominios amino terminales del dímero de la subunidad  $\alpha$ . La subunidad  $\omega$  se encuentra

envolviendo al dominio carboxilo terminal de la subunidad  $\beta'$  (Borukhov y Lee, 2005).

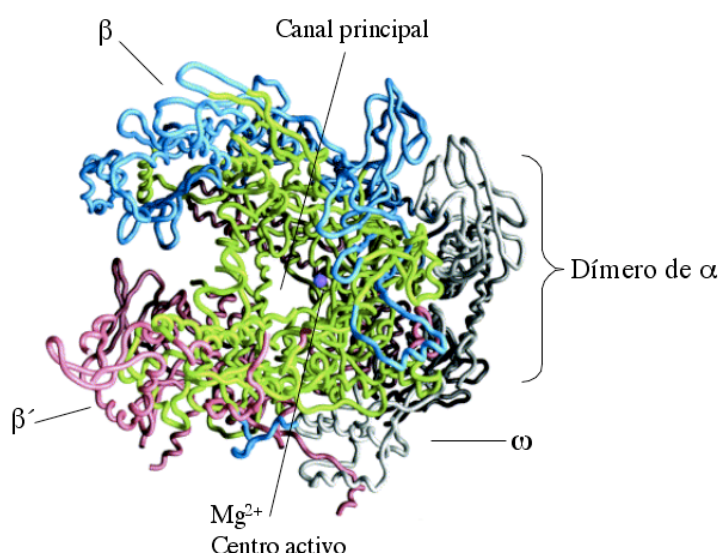


Figura 1: Diagrama de la estructura del *core* de la RNAP de *Taq* (Darst, 2001). En verde se muestran las regiones conservadas evolutivamente desde procariotas a eucariotas. En azul la subunidad  $\beta$  y en rosa la  $\beta'$ . El dímero de los dominios NTD de las subunidades  $\alpha$  se representa en gris oscuro, y en gris claro la subunidad  $\omega$ . Se puede apreciar que tiene aspecto de pinza de cangrejo, con el  $Mg^{2+}$  (morado) en el centro activo de la enzima.

El espacio entre las pinzas está atravesado por tres canales: el “canal principal” tiene un diámetro entre 20-27 Å y es donde se acomodan el DNA de doble cadena y el híbrido DNA/RNA. Existen otros dos canales menores con un diámetro de 10-12 Å, que derivan del canal principal y son el “canal secundario” y el “canal de salida del RNA”. El “canal secundario” tiene dos funciones; por un lado, es el lugar de entrada de los NTPs para acceder al centro activo y por otro, acomoda al



extremo 3' del RNA cuando se produce la parada y el retroceso de la enzima. El otro canal menor es por donde sale el RNA de banda sencilla recién sintetizado (Darst, 2001; Komissarova y Kashlev, 1997; Korzheva *et al.*, 2000; Severinov, 2000).

Una  $\alpha$ -hélice de la subunidad  $\beta'$  conecta las pinzas en la zona próxima al centro activo y un bucle, también de  $\beta'$ , forma la pared del canal secundario; las paredes del canal de salida del RNA están formadas por diferentes regiones de  $\beta$  y  $\beta'$ . El centro activo se encuentra localizado en una región altamente conservada de la subunidad  $\beta'$  y está formado por una triada catalítica de aspárticos (NADFDGD) que mantienen al ión magnesio, esencial para la función, localizado en la pared del canal principal (Murakami *et al.*, 2002; Westover *et al.*, 2004; Zhang *et al.*, 1999).

**Las subunidades  $\alpha$**  (de aproximadamente 37 KDa) forman un dímero que es el iniciador en el proceso de ensamblaje de la RNAP; contienen dos dominios independientes unidos por una región flexible de, al menos, 13 aminoácidos (aa) (Blatter *et al.*, 1994; Jeon *et al.*, 1997; Negishi *et al.*, 1995). El dominio amino terminal (NTD; aa del 1-233, en *E. coli*) es el responsable de la dimerización y de las interacciones con el *core*, en concreto con  $\beta$  y  $\beta'$  (Barker *et al.*, 2001). Recientemente se ha demostrado que puede ser, también, un dominio susceptible para la regulación de la transcripción (Browning y Busby, 2004). El dominio carboxilo terminal (CTD; aa del 249-329, en *E. coli*) contiene un determinante débil de dimerización, determinantes para la interacción con el DNA, tanto con secuencias específicas (elementos UP) como con secuencias inespecíficas y, además, posee determinantes para la interacción con la subunidad sigma, activadores, represores, factores de elongación y de terminación. Los  $\alpha$ -CTDs son los dominios del *core* diana por excelencia para la interacción con factores de regulación (Busby y Ebright, 1999; Murakami *et al.*, 1996; Savery *et al.*, 2002; Savery *et al.*, 1998). El conector que une ambos dominios, debido a su flexibilidad, permite a la subunidad  $\alpha$ -CTD localizarse en diferentes posiciones respecto al resto de la enzima y así formar parte de distintos tipos de complejos (Blatter *et al.*, 1994; Busby y Ebright, 1994; Ebright y Busby, 1995). Las subunidades  $\alpha$  presentan alta afinidad por unas secuencias de DNA llamadas elementos UP; son secuencias muy ricas en pares de bases A/T con el surco menor estrechado, que es precisamente por donde interacciona el dominio  $\alpha$ -CTD (Gourse *et al.*, 2000; Savery *et al.*, 1995).

Se resolvió la estructura del dominio  $\alpha$ -CTD de la RNAP de *E. coli* unido al DNA (Benoff *et al.*, 2002) y se pudieron determinar dos razones estructurales que justifican su preferencia por secuencias ricas en A/T: en primer lugar, la Arg265 interacciona con la espina de hidratación que se forma en el surco menor de esta secuencia de DNA, estableciendo puentes de hidrógeno, mediados por agua, con las bases del DNA, y en segundo lugar, la Arg265 contacta con una red de interacciones que se producen a lo largo de la estructura comprimida, típica de las secuencias ricas en A/T, debido al estrechamiento del surco menor (Katahira *et al.*, 1990; Kopka *et al.*, 1983; MacDonald *et al.*, 2001). Existen dos motivos *HhH* (Helix-hairpin-Helix) en el  $\alpha$ -CTD implicados en el reconocimiento del DNA y, además del residuo Arg265 crítico para la unión, otros seis aminoácidos, de forma directa o mediada por agua, interaccionan con el elemento UP (Val264, Ser299, Asn268, Asn294, Gly296, Lys298 y Arg265) formando el llamado “Determinante 265” (Benoff *et al.*, 2002; Ross *et al.*, 2003).

**La subunidad  $\omega$**  es un polipéptido de 91 aa que no tiene un papel directo en la transcripción y que parece funcionar como chaperona para ayudar al plegamiento de la subunidad  $\beta'$  (Mathew y Chatterji, 2006; Minakhin *et al.*, 2001), aunque también se ha propuesto que influye en la respuesta celular ante condiciones ambientales restrictivas y

severas (Vrentas *et al.*, 2005). Además, *B. subtilis* tiene una subunidad  $\delta$  que modifica la velocidad de síntesis del RNA y se ha propuesto que tiene también, un papel en la asociación de la RNAP al DNA (Juang y Helmann, 1994a; López de Saro *et al.*, 1999).

**Subunidad  $\sigma$ :** Las bacterias poseen más de un factor  $\sigma$ . La utilización de subunidades  $\sigma$  alternativas capacita a la RNAP para reconocer diferentes clases de promotores (Lonetto *et al.*, 1992; Wosten, 1998). Esto permite a los organismos, a través de una transcripción específica y selectiva, llevar a cabo los procesos requeridos para su óptimo crecimiento celular, así como responder rápidamente a cambios producidos en el entorno que los rodea. Existe un factor  $\sigma$  vegetativo o primario ( $\sigma^{70}$  en *E. coli* y  $\sigma^A$  en *B. subtilis*), que es el responsable de la expresión de la mayoría de los genes durante el crecimiento exponencial de las células. El resto de factores  $\sigma$  se unen al *core* de la RNAP para conducir la expresión de genes requeridos en determinados momentos o situaciones especiales. Como resultado se obtienen *holoenzimas* que reconocen promotores concretos ante situaciones determinadas.

La subunidad  $\sigma$  tiene tres funciones principales: i) Reconocer secuencias específicas del promotor, ii) posicionar a la RNAP holoenzima en el promotor diana y iii) facilitar la separación de las cadenas de DNA cerca del sitio de inicio de la transcripción (Wosten, 1998). Además, la subunidad  $\sigma$  es una buena diana para muchos reguladores de la transcripción (Dove *et al.*, 2003).

Las subunidades  $\sigma$  vegetativas ( $\sigma^{70}$  y similares) son las más abundantes y poseen cuatro dominios estructurales,  $\sigma^{1.1}$ ,  $\sigma^2$ ,  $\sigma^3$  y  $\sigma^4$ , unidos por brazos flexibles; los dominios contienen, respectivamente, las regiones 1.1, 1.2-2.4, 3.0-3.1 y 4.1-4.2, de secuencias altamente conservadas; la región 3.2 conecta los dominios 3 y 4 (Gruber y Gross, 2003; Lonetto *et al.*, 1992).

En la estructura del *holoenzima* de *Tht*, la subunidad  $\sigma$  tiene forma de V y está localizada entre las subunidades  $\beta$  y  $\beta'$ . La mayoría de las interacciones entre  $\sigma$  y el *core* son relativamente débiles y se distribuyen a lo largo de una amplia superficie; prácticamente todas ellas, tienen lugar con las subunidades  $\beta$  y  $\beta'$  (Borukhov y Lee, 2005; Borukhov y Nudler, 2003; Gruber *et al.*, 2001; Owens *et al.*, 1998; Sharp *et al.*, 1999). Existen también interacciones entre la subunidad  $\sigma^4$  y el dominio  $\alpha$ -CTD, que pueden estar o no mediadas por activadores de la transcripción (Chen *et al.*, 2003; Ross *et al.*, 2003).

**$\sigma^{1.1}$ :** La secuencia de este dominio no está muy conservada, pero sí su carácter ácido. Se ha demostrado que esta región inhibe la unión de  $\sigma$  al promotor cuando está libre; tras la asociación con el *core* para formar la *holoenzima*, se produce un cambio conformacional que hace que  $\sigma^{1.1}$  se desplace y se permita la interacción con el promotor. Se ha sugerido, también, que podría participar favoreciendo la formación del complejo abierto (Bowers y Dombroski, 1999; Camarero *et al.*, 2002; Mekler *et al.*, 2002; Vuthoori *et al.*, 2001; Wilson y Dombroski, 1997).

**$\sigma^2$ :** La región 2.4 es la responsable del reconocimiento de una de las secuencias conservadas del promotor, el hexámero -10. La región 2.3 es esencial para la separación de las cadenas de DNA y para la estabilización de la burbuja de transcripción al inicio del proceso (Aiyar *et al.*, 1994; Juang y Helmann, 1994b).

**$\sigma^3$ :** Los aminoácidos de la región 3.0, además de estabilizar el complejo abierto, están implicados en el reconocimiento de secuencias del promotor (Barne *et al.*, 1997). El conector de la región 3.2 parece que está implicado en la unión correcta del nucleótido iniciador, aunque no se sabe si de forma directa o indirecta (Campbell *et al.*, 2002).

**$\sigma_4$ :** Este dominio tiene forma de C con un bolsillo cóncavo recubierto por residuos hidrofóbicos de la región 4.1 implicados en la interacción con el *core*. La región 4.2 es la responsable del reconocimiento de otro de los elementos conservados del promotor, el hexámero  $-35$ . La unión a esta secuencia produce una curvatura que altera la trayectoria del DNA (Campbell *et al.*, 2002; Murakami *et al.*, 2002; Vassilyev *et al.*, 2002). Esta modificación conformacional facilita las interacciones del  $\alpha$ -CTD con el DNA y con activadores que se unan por debajo de este elemento  $-35$  (Murakami y Darst, 2003; Sharp *et al.*, 1999).

Existe mucha controversia con respecto al momento en el que la subunidad  $\sigma$  se disocia del holoenzima y si ésto ocurre en realidad; el modelo que parece explicar mejor las observaciones más recientes implica la disociación estocástica del factor  $\sigma$ , es decir, que puede disociarse en momentos muy distintos del proceso de transcripción y que ocurre de forma fortuita, pudiendo existir, incluso, reasociaciones de  $\sigma$  en determinados momentos a lo largo del proceso de elongación (Mooney *et al.*, 2005).

## 1.2) Promotores

Los promotores son señales en el DNA que indican dónde debe comenzar la transcripción, cuánto de fuerte debe ser y cuándo debe ocurrir, ya que, de sus secuencias y de las secuencias próximas va a depender la interacción de la RNAP y la unión de activadores y/o represores que regulen el proceso.

El inicio de la transcripción requiere la interacción de la RNAP con el DNA del promotor. Se han identificado cuatro elementos diferentes en la secuencia del DNA, que son los responsables del reconocimiento en promotores dependientes de factores  $\sigma$  primarios o vegetativos (figura 2).

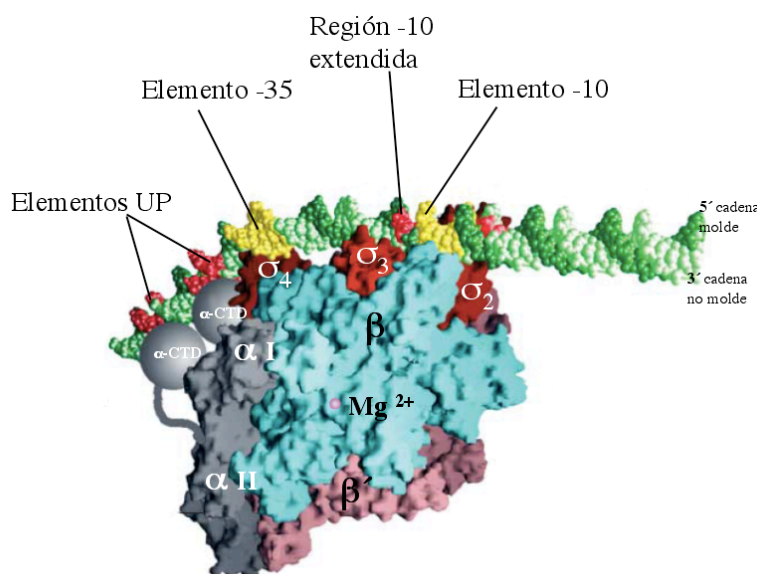


Figura 2: Modelo, basado en los estudios cristalográficos, del holoenzima de la RNAP unido al promotor. Adaptado de (Browning y Busby, 2004). Sobre la doble hélice de DNA, en verde, se muestran los elementos  $-10$  y  $-35$  en amarillo, y en rojo la región  $-10$  extendida y los elementos UP. Se presentan los dominios de la subunidad  $\sigma$  (en rojo) implicados en el reconocimiento del promotor y las subunidades del *core*  $\beta$  en azul y  $\beta'$  en rosa. Las subunidades  $\alpha$ , en gris, se representan interactuando con los elementos UP.

Los dos principales elementos del promotor, el **hexámero  $-10$**  y el **hexámero  $-35$** , tienen las siguientes secuencias consenso y posiciones respecto al punto de inicio de la transcripción: *el elemento  $-10$* :  $^{-12}$  TATAAT  $^{-7}$  y *el elemento  $-35$* :  $^{-35}$  TTGACA  $^{-30}$ ; se encuentran separados por una secuencia inespecífica de 16-18 pares de bases, siendo 17 pb la distancia óptima (Lisser y Margalit, 1993; McClure, 1985).

El elemento  $-10$  es reconocido por la región 2.4 de la subunidad  $\sigma$  de la RNAP, mientras el elemento  $-35$  lo es por la región 4.2. La subunidad  $\sigma$  es la encargada de

elegir el punto de inicio de la transcripción y tiene lugar a una distancia entre 6 y 8 nucleótidos de la región  $-10$  y preferiblemente en purinas (Lisser y Margalit, 1993).

Los otros dos elementos reconocidos son la región  **$-10$  extendida** y el **elemento UP**. La *región  $-10$  extendida* es un dinucleótido  $5'TG\ 3'$  localizado a continuación del hexámero  $-10$ , en las posiciones  $-14$  y  $-15$  y es reconocido por el dominio 3 de la subunidad  $\sigma$  (Belyaeva *et al.*, 1993; Keilty y Rosenberg, 1987; Kumar *et al.*, 1993; Sanderson *et al.*, 2003). Este elemento  $-10$  extendido sustituye en *E. coli*, en ocasiones, al elemento  $-35$ , mientras que en muchos promotores de *B. subtilis* están presentes ambos elementos. El *elemento UP* es una secuencia de aproximadamente 20 pares de bases (pb) localizada más arriba del hexámero  $-35$  (de la posición  $-40$  a la  $-60$ ) y es reconocida por el dominio CTD de la subunidad  $\alpha$  de la RNAP. Se caracteriza por ser una región rica en bases A/T, y posee dos subsitios, uno para cada una de las subunidades  $\alpha$ . Se ha visto que la interacción con el sitio proximal ejerce un mayor efecto sobre el reconocimiento y activación del promotor, probablemente porque se producen también interacciones del  $\alpha$ -CTD con la subunidad  $\sigma$  unida en  $-35$  (Aiyar *et al.*, 1998; Benoff *et al.*, 2002; Busby y Ebright, 1999; Gourse *et al.*, 2000; Ross *et al.*, 2001; Ross *et al.*, 1993).

Estos cuatro elementos de reconocimiento establecen la unión inicial de la RNAP al promotor, están involucrados en su interacción directa y específica y, por lo tanto, en la formación del complejo cerrado. La contribución relativa de cada uno de estos elementos difiere de un promotor a otro. Cuanto más se parezcan las secuencias de reconocimiento a las secuencias consenso, más fuerte es la interacción de la RNAP con el promotor.

A veces existen secuencias adicionales de DNA localizadas en y alrededor del promotor, que pueden compensar un elemento  $-10$  ó  $-35$  débil y afectar a la fuerza general del promotor; se demostró que estos elementos auxiliares del promotor, como la región discriminatoria (entre  $-10$  y  $+1$ ) y la primera región transcrita, afectan a la velocidad de formación del complejo abierto, a la eficiencia de separación de las cadenas, a la iniciación abortiva y al escape del promotor (Ellinger *et al.*, 1994; Liu *et al.*, 2004; Pérez-Martín y Espinosa, 1994).

Muchos promotores pueden funcionar en ausencia de reguladores mientras que en otros casos, debido a que sus elementos no son muy cercanos al consenso o a que sus espaciadores no poseen la distancia óptima, requieren la intervención de un activador; de hecho, hay promotores programados para responder a señales reguladoras que van a modular su actividad aumentando o disminuyendo la velocidad de una iniciación productiva.

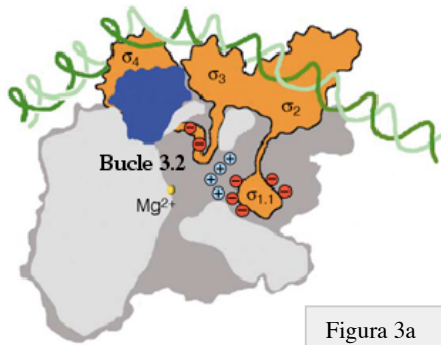
### 1.3) El proceso de transcripción. La iniciación.

Un ciclo de transcripción, como se ha comentado anteriormente, tiene lugar a través de tres estadios: iniciación, elongación y terminación.

Durante la iniciación, la RNAP, como *holoenzima*, a través de la interacción con los elementos  $-10$  y  $-35$ , se une específicamente al promotor y da lugar a la formación del complejo cerrado (cc). Posteriormente, en un proceso que involucra varios estados intermedios, se produce una isomerización que da lugar a la formación de un complejo abierto estable (oc); en este paso, el DNA de doble cadena se deshíbrida en la región próxima a la posición  $-10$  (desde la  $-12$  a la  $+3$ ) y se forma la burbuja de transcripción (Tomsic *et al.*, 2001; Tsujikawa *et al.*, 2002). En presencia de los NTPs, la RNAP empieza a sintetizar y liberar pequeños RNAs (de 2-12 nts), llamándose este proceso iniciación abortiva; la RNAP no abandona el promotor, y tras varias rondas de iniciación abortiva, el complejo de iniciación entra en el estadio de elongación. Esta

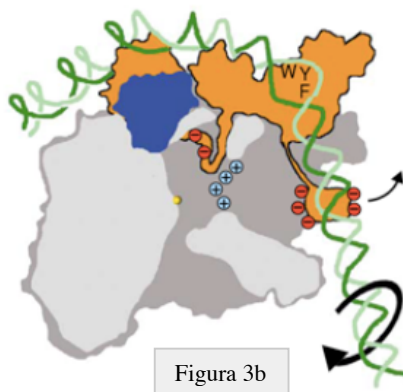
transición, que implica el escape del promotor, está marcada por un cambio conformacional y la formación de un complejo ternario de elongación altamente procesivo. La elongación continúa hasta que la RNAP transcribe una señal de terminación codificada en el DNA y que consiste en secuencias con simetría invertida, tras lo que se produce una disociación irreversible del complejo formado por el *core*, el DNA y el RNA. Puesto que el paso más fuertemente regulado del proceso es la iniciación, se describe a continuación con más detalle.

Iniciación de la transcripción (figura 3) adaptado de (Murakami y Darst, 2003):



#### Complejo cerrado (cc):

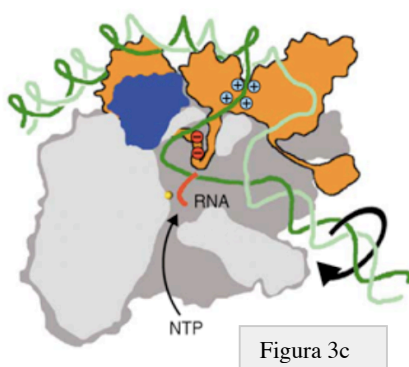
En primer lugar se produce la interacción de las regiones 2.4 y 4.2 de la subunidad  $\sigma$  con los elementos  $-10$  y  $-35$  del DNA, reconociéndose así el promotor. Como se puede ver en la figura 3a, el bucle 3.2 de  $\sigma$  se introduce en el canal donde está el centro activo de la enzima, representado por el ión  $Mg^{2+}$ ; el dominio 1.1, se encuentra también en el canal del centro activo, en concreto, en el canal por donde sale el DNA de doble banda.



#### Pasos intermedios:

La región 2.3 de  $\sigma$  contiene unos aminoácidos aromáticos muy conservados (WYF) que interaccionan con la cadena complementaria del elemento  $-10$ , aprovechando la “respiración” típica de las secuencias ricas en A/T. De este modo, se secuestra la cadena y como resultado se produce la estabilización inicial de la burbuja de transcripción (figura 3b). Esta separación de las bandas provoca que el DNA se pueda desenrollar, adquiriendo una cierta flexibilidad que le permite girar para alcanzar el centro activo. La entrada del

DNA en el canal, entre las posiciones  $+5$  y  $+12$ , debe ir acompañada de la salida del dominio 1.1, lo cual ocurre por un mecanismo desconocido.



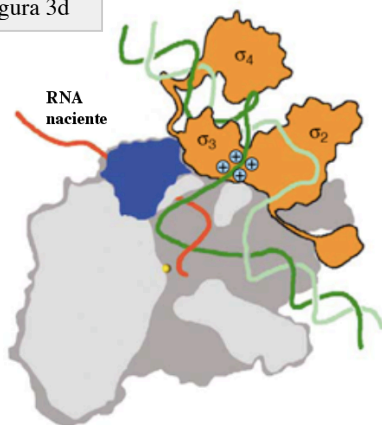
#### Complejo abierto (oc) e iniciación abortiva:

En el complejo abierto final, el DNA en la burbuja de transcripción está deshibridado hasta más allá del punto de inicio ( $+1$ ). La región  $\sigma 1.2$  establece interacciones con la cadena complementaria, evitando un colapso de la burbuja de transcripción (Shanil, 2006). La cadena molde del DNA es dirigida al centro activo a través de una abertura cargada positivamente. El DNA de doble banda por delante del sitio de inicio ( $+5$  a  $+12$ ) se encuentra dentro de un túnel proteico entre  $\beta$  y  $\beta'$ ; este DNA es muy importante para la

estabilidad del complejo. Por el canal secundario entran los NTPs, y la RNAP empieza a sintetizar RNAs de pequeña longitud que se encuentran con el bucle 3.2 en su canal de salida, dando lugar a transcritos abortivos que probablemente salgan a través del canal secundario (figura 3c).

Cuando la cadena de RNA se elonga más de 12 nucleótidos llena la cavidad del híbrido DNA-RNA y desplaza el bucle de 3.2 alcanzándose así el final de la iniciación abortiva.

Figura 3d



#### Escape del promotor:

En las fases iniciales del escape del promotor tiene lugar la desestabilización de las interacciones entre  $\sigma_4$  y  $\beta$ . Esto podría, a su vez, desestabilizar las interacciones entre  $\sigma_4$  y el elemento  $-35$  permitiendo que la RNAP salga del promotor (figura 3d). La transición a la elongación no requiere la separación completa de  $\sigma$ ; se ha visto que la interacción con  $\sigma_2$  y  $\sigma_3$  no es incompatible en el complejo ternario de elongación.

Finalmente, y de acuerdo con la teoría estocástica mencionada anteriormente, tiene lugar la disociación completa de  $\sigma$  de la *holoenzima* y es el *core* de la RNAP el que lleva a cabo la elongación del nuevo RNA hasta su terminación (figura 3e).



Figura 3e

### **1.4) Modulación de la transcripción**

En cualquier tipo celular debe existir una rigurosa regulación de la expresión génica, para mantener sus funciones basales y para poder adaptarse rápida y eficazmente ante cualquier cambio que se produzca en su entorno próximo. La expresión génica puede ser regulada en muchos puntos del proceso, a nivel de la transcripción, en la traducción de proteínas, mediante modificaciones postraduccionales o, incluso, por degradación proteica. Uno de los procesos más ampliamente estudiado es la *regulación de la transcripción* que, a su vez, puede ser regulada a nivel de la iniciación, de la elongación o de la terminación.

La iniciación de la transcripción es el paso más susceptible de regulación y también el más conocido y estudiado. Podemos clasificar en cuatro los mecanismos moleculares de regulación del inicio de la transcripción, que aseguran una adecuada distribución de la RNAP en una bacteria que posee muchos promotores compitiendo por ella.

#### 1.4.1) Secuencias de los promotores

En el genoma de una célula bacteriana existen más de 2000 promotores que son reconocidos por la RNAP (Salgado *et al.*, 2001). Cuanto más se parezcan las secuencias de un promotor a las consenso, más eficiente es el reconocimiento de dicho promotor y la consecuente formación del complejo cerrado. En cambio, la realidad es que apenas existen promotores que posean todas las secuencias consenso y por lo tanto, la mayoría de los promotores funcionan por debajo de su nivel óptimo. Este hecho es, en sí mismo, una forma de regular la transcripción, ya que los hace más susceptibles de ser activados cuando la ocasión lo requiera (Browning y Busby, 2004).

#### 1.4.2) Regulación por factores $\sigma$ alternativos:

Además del factor  $\sigma$  vegetativo, existen otros factores  $\sigma$  que se producen en respuesta a situaciones específicas de estrés, que al irse acumulando generan una

situación de competición con la subunidad  $\sigma$  vegetativa por la RNAP (Ishihama, 2000). De modo que son capaces de interaccionar con el *core* de la RNAP y conducir la expresión de promotores con secuencias particulares. La actividad de los factores  $\sigma$  está rigurosamente controlada; además de la regulación de su expresión, también se puede modular su actividad mediante factores antisigma. Estos factores funcionan secuestrando a la subunidad  $\sigma$  correspondiente y evitando que esté libre para interaccionar con el *core*, o bien controlando su proteólisis; los factores antisigma, además, pueden influir en la transcripción de promotores por medio de la regulación de su propia síntesis y degradación (Campbell *et al.*, 2002; Campbell *et al.*, 2003; Helmann, 1999; Hughes y Mathee, 1998; Sorenson *et al.*, 2004).

#### 1.4.3) Pequeños ligandos

Los pequeños ligandos permiten un mecanismo alternativo por el cual la RNAP puede responder rápida y eficientemente a los cambios en el entorno de la célula. El ejemplo mejor y más estudiado es el de la guanosina 3',5'-bifosfato (ppGpp). Este ligando es sintetizado cuando la disponibilidad de aminoácidos está restringida hasta el punto de que la traducción se ve limitada. El ligando ppGpp desestabiliza el complejo abierto de transcripción en los promotores que sintetizan proteínas implicadas en la maquinaria de traducción. De esta manera, ppGpp inhibe la expresión de la maquinaria de traducción en respuesta a una repentina falta de nutrientes (Chatterji y Ojha, 2001).

#### 1.4.4) Factores de transcripción

Los factores de transcripción son proteínas reguladoras que afectan al inicio del proceso, bien favoreciéndolo, como hacen los llamados activadores, o bien desfavoreciéndolo como hacen los represores. Tanto unos como otros pueden funcionar por mecanismos muy diferentes en los que se pueden encontrar involucradas una o más moléculas, y estos mecanismos pueden, a su vez, ser muy sencillos o realmente complejos. En general, cuando un factor de transcripción se une a un promotor puede activarlo o reprimirlo, pero existen también factores de transcripción que en función del promotor del que se trate pueden funcionar como activadores o como represores (Browning y Busby, 2004; Pérez-Rueda y Collado-Vides, 2000).

La actividad de un factor transcripcional puede ser modulada *modificando su afinidad de unión al DNA*. Esta afinidad puede, a su vez, estar regulada por medio de ligandos que se generan en respuesta a una determinada situación, o por medio de modificaciones covalentes como la fosforilación, que suele depender de una cascada de reacciones en cadena desde el exterior celular (Stock *et al.*, 2000). También la *disponibilidad de un factor transcripcional* puede modificar su actividad global y ésta depende tanto de su expresión como de su degradación (Dempfle, 1996). Por último, la actividad de estos factores también puede ser regulada por medio de *otras proteínas* que interaccionando con ellos les impidan o les ayuden a llevar a cabo su función, ya sea activadora o represora.

##### 1.4.4.a) Mecanismos Simples de Regulación mediados por factores de transcripción (figura 4):

**A) Activación:** existen tres mecanismos básicos para regular, de una forma sencilla, la activación de la transcripción (Browning y Busby, 2004).

##### a) Activación de Clase I (figura 4.A.a):

En promotores de clase I, los reguladores se unen al DNA en posiciones bastante alejadas del punto de inicio de la transcripción, centradas en -61, -71, -81 ó -91. El



factor de transcripción unido en estas posiciones interacciona con la subunidad  $\alpha$ -CTD de la RNAP reclutando a la enzima en el promotor (Ebright, 1993). La flexibilidad que posee el brazo que une ambos dominios de la subunidad  $\alpha$  permite que el activador pueda unirse en muchas posiciones diferentes con el mismo resultado de activación del promotor.

b) Activación de Clase II (figura 4.A.b):

En los promotores de clase II, el lugar donde se unen los reguladores está centrado en la posición  $-42$ , solapando con la región  $-35$  del promotor y en la mayoría de los casos ejercen su función interaccionando con la región 4 de la subunidad  $\sigma$  (Dove *et al.*, 2003). La posición donde se unen los reguladores en los promotores de clase II, al contrario de lo que ocurre en los de clase I, no puede variar porque la subunidad  $\sigma$  tiene una posición fija en el promotor. Estos activadores, en ocasiones, pueden interaccionar con otras subunidades de la RNAP como puede ser el dominio NTD de la subunidad  $\alpha$  (Savery *et al.*, 1998) o el  $\alpha$ -CTD que se uniría por detrás del sitio de unión del activador (Busby y Ebright, 1997). En este tipo de promotores, aunque también se produce una estabilización de la RNAP en el promotor, la función activadora del regulador es debida, fundamentalmente, a que su interacción con  $\sigma$  produce cambios conformacionales en la RNAP que favorecen la isomerización en la formación del complejo abierto (Busby y Ebright, 1997).

c) Mecanismo de Activación mediado por Cambios Conformacionales en el DNA (figura 4.A.c):

La proteína reguladora produce un cambio conformacional en el DNA del promotor diana permitiendo así la interacción de la RNAP con el elemento  $-10$  y/o con el elemento  $-35$ . Un ejemplo sería el caso en el que la distancia entre ambos elementos del promotor no fuese la óptima, la unión del activador los acercaría permitiendo la interacción de la RNAP. En general se requiere la unión del activador en/o muy cerca de los elementos del promotor, aunque también hay algún caso descrito de activación a distancia (Sheridan *et al.*, 1998; Sheridan *et al.*, 2001).

**B) Represión:** los procesos básicos de represión simple mediada por factores transcripcionales se pueden resumir en tres:

a) Represión por impedimento estérico (figura 4.B.a):

En estos casos, el represor se une al DNA del promotor interfiriendo con la unión de la RNAP en el mismo. En unos casos puede estar totalmente impedida la unión de la enzima mientras que en otros, aunque se permite la unión, se impiden pasos posteriores al reclutamiento de la RNAP (Muller-Hill, 1998).

b) Represión por reestructuración del DNA (figura 4.B.b):

Los represores se unen a distancia del promotor, y por interacciones entre ellos se forma una estructura en el DNA que, o bien envuelve al promotor y lo hace inaccesible a la RNAP, o bien provoca una curvatura que hace difícil la estabilización de la enzima.

c) Antiactivadores (figura 4.B.c):

Impiden la activación de un promotor dependiente de activador. El represor interaccionando con el activador lo secuestra o lo modifica para evitar la activación (Shin *et al.*, 2001; Valentin-Hansen *et al.*, 1996).



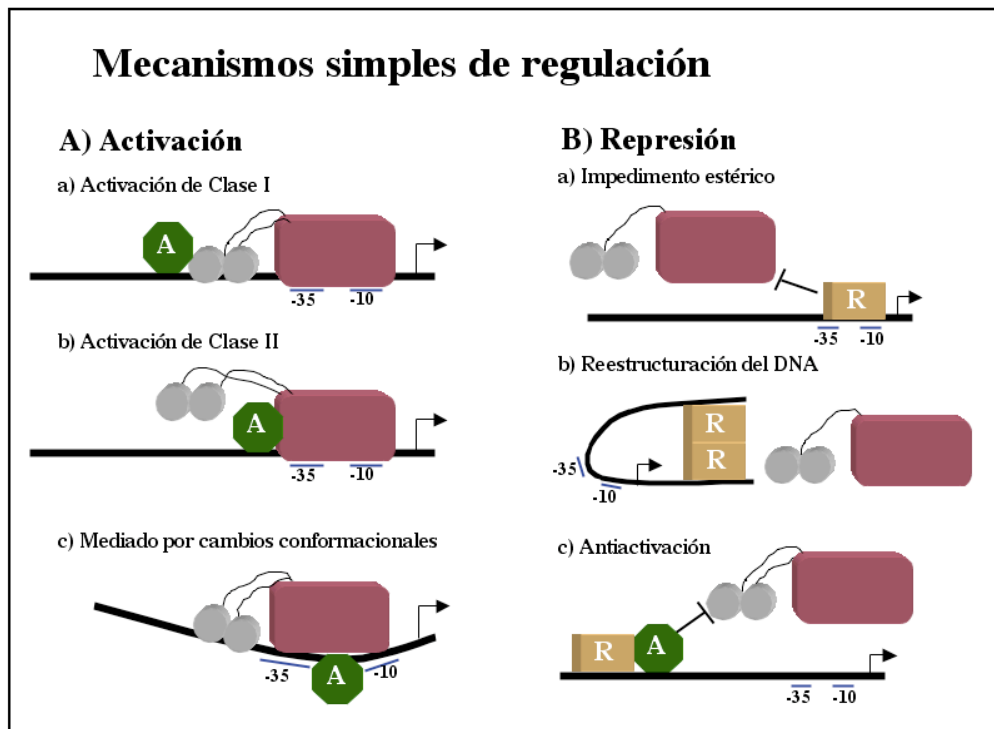


Figura 4: Dibujo esquematizado de los diferentes mecanismos sencillos de regulación mediados por factores de transcripción. Se muestran los activadores en verde, los represores en naranja, la RNAP en granate, y sus dominios  $\alpha$ -CTD, en gris.

Los promotores cuya activación es dependiente de factores de transcripción tienen un funcionamiento, generalmente, muy complicado; de hecho, suelen estar co-regulados por un represor, por un activador o por ambos. Hay que tener en cuenta que las células necesitan acoplar la expresión de genes individuales a todas las señales ambientales que recibe y que son interpretadas a través de los diferentes factores de transcripción. Los mecanismos usados para ello no están todavía del todo claros.

#### 1.4.4 b) Mecanismos Complejos de Regulación mediados por factores de transcripción (figura 5):

Implican mecanismos por los cuales múltiples reguladores pueden unirse a un promotor para integrar diferentes estímulos y dar lugar a una respuesta apropiada (Barnard *et al.*, 2004).

##### **A) Codependencia:**

###### a) Contactos independientes con la RNAP:

La activación del promotor depende de que dos activadores diferentes interactúen con la RNAP de forma independiente. Este tipo de regulación se puede dar de dos formas:

1) Promotores Clase II/ Clase I (figura 5.A.a.1): en estos promotores, uno de los activadores interactúa con la RNAP por medio de un mecanismo de clase I, es decir, interactuando con el dominio CTD de una de las subunidades  $\alpha$  de la RNAP, mientras que el otro activador lo hace por medio de un mecanismo de clase II, uniéndose al DNA en una región muy próxima al elemento  $-35$ , de modo que existe interacción con la subunidad  $\sigma$  de la RNAP. En el promotor *proP* P2 de *E. coli*, Fis

funciona como un activador de clase II y CRP como un activador de clase I (Busby y Ebright, 1999).

2) Promotores ClaseI/ClaseI (figura 5.A.a.2): en estos promotores, cada activador interacciona con el dominio CTD de una subunidad  $\alpha$ , provocando la estabilización de la RNAP en el promotor; este es el mecanismo que utilizan dos dímeros de CRP en el promotor *acs* P2 de *E. coli* (Beatty *et al.*, 2003).

b) Codependencia de la activación debida a la unión cooperativa de los activadores (figura 5.A.b):

Se necesita la unión de dos activadores para que tenga lugar la activación; en este caso, la unión de un activador al DNA requiere la unión del otro y ésta es cooperativa, de modo que uno sin el otro no dan lugar a la activación del promotor. Para que el promotor *melAB* de *E. coli* sea activado se requiere la interacción de MelR con la RNAP; para que esto tenga lugar, se necesita la unión previa de CRP entre los sitios de unión de MelR (Grainger *et al.*, 2004; Wade *et al.*, 2001).

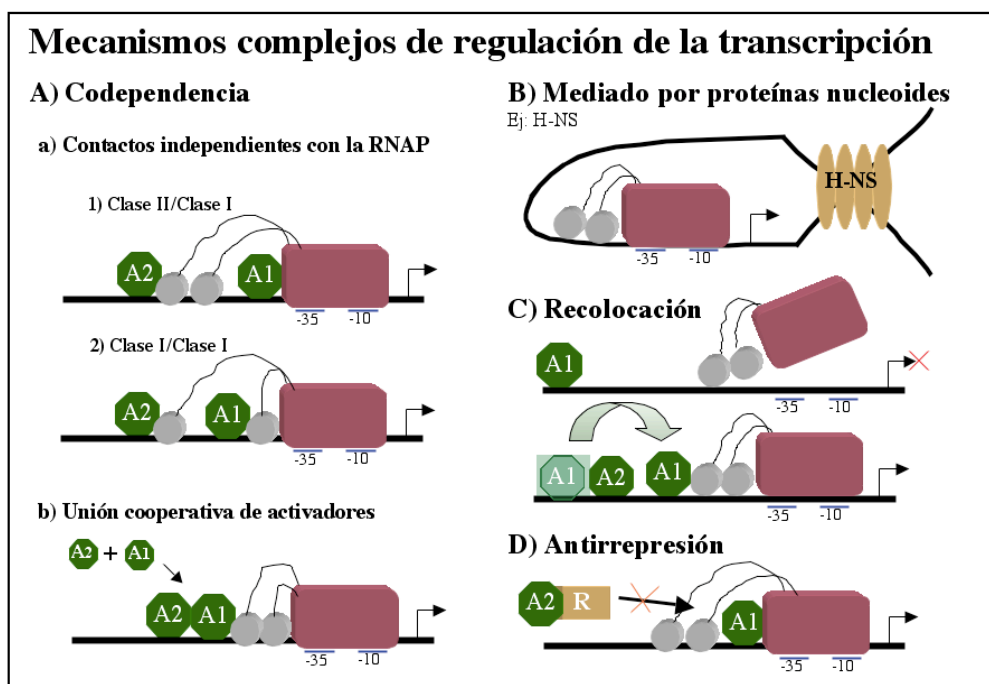


Figura 5: Dibujo esquematizado de los diferentes mecanismos complejos de regulación de la transcripción mediados por factores de transcripción. Se muestran cuatro ejemplos tipo, que implican la intervención de más de una proteína. Como en la figura anterior, en verde se representan los activadores, en naranja lo represores y la RNAP en granate y gris.

### B) Regulación mediada por proteínas nucleoides ó proteínas tipo histona:

Las bacterias poseen unas proteínas con similitud a las histonas eucariotas. Esta similitud no es debida a la secuencia de aminoácidos, sino por su capacidad para unirse al DNA, su baja masa molecular, su número de copias y su carga electrostática. Estas proteínas están involucradas en la organización del nucleóide como hacen las histonas en la cromatina. Muchas de estas proteínas bacterianas influyen en la expresión de genes y en otras funciones en las que está implicado el DNA, como pueden ser la recombinación y la replicación. Las mejor caracterizadas son Fis, H-NS, IHF y HU, de *E. coli* (Dorman y Deighan, 2003).

**HU:** Es una proteína con un papel básicamente estructural que se une como dímero al DNA curvándolo. Se conocen muchos casos en los que HU controla la

actividad de promotores; uno de esos casos es en el promotor *gal* P2 de *E. coli* donde HU colabora estabilizando la formación de un bucle en el DNA por la proteína tetramérica GalR, que tiene como objetivo la represión de la transcripción (Semsey *et al.*, 2002; Kar y Adhya, 2001).

**H-NS:** Esta proteína es generalmente un represor en el control de la expresión de genes. Aproximadamente el 5% de los genes de *E. coli* responden a su presencia o ausencia. Los pocos casos en los que afecta positivamente a la expresión involucran efectos indirectos o postraduccionales. H-NS se une a regiones intrínsecamente curvadas en el DNA del promotor que va a ser reprimido; existen varios sitios de unión de la proteína y se forma un bucle donde es atrapada la RNAP. Es lo que ocurre en el promotor *rrnB* P1 del ribosoma en *E. coli* (Dame *et al.*, 2001; Schroder y Wagner, 2000) (figura 5.B).

**Fis:** Desarrolla un papel muy importante en la regulación de la expresión génica en *E. coli*, y lo hace por mecanismos muy diversos; es capaz de contrarrestar la represión provocada por H-NS en los promotores *rrnB*, *virF* y *hns*; por lo tanto, es en sí mismo un activador de la transcripción de esos promotores (Falconi *et al.*, 2001). Es también represor de la transcripción y regula el promotor que controla la propia expresión de su gen *fis* (Walker *et al.*, 1999). Además, reprime otros muchos promotores por medio de diferentes mecanismos como pueden ser la oclusión del promotor en *gyrA* o el impedimento de la isomerización a complejo abierto en *gyrB* (Schneider *et al.*, 1999).

#### **C) Mecanismo de Recolocación (figura 5.C):**

En este caso, la activación del promotor depende sólo de la unión de un activador. En ausencia de un segundo activador, el primero se coloca erróneamente para la activación del promotor, de modo que el papel del segundo activador es sencillamente posicionar al primero en el lugar adecuado para que la activación tenga lugar. Esto es lo que ocurre en el promotor *malK* de *E. coli* donde CRP es el responsable del correcto posicionamiento del activador MalT. También puede darse porque un segundo activador produzca un cambio conformacional en el DNA que permita las interacciones de un primer activador con la RNAP, sin modificar el posicionamiento del primer activador (Richet *et al.*, 1991).

#### **D) Antirrepresión (figura 5.D):**

En este mecanismo, un activador que por sí mismo da lugar a la activación del promotor, se encuentra inhibido por un represor; es entonces necesario un segundo activador capaz de interactuar con el represor, liberando al primer activador del represor y dando lugar a la activación. Este represor suele ser una proteína nucleoidea. En muchas ocasiones las proteínas nucleoides se encuentran formando complejos nucleoproteicos que reprimen la transcripción de un promotor. La llegada de un activador puede modificar o desplazar la unión de estas proteínas y dar como resultado la activación del promotor, como ocurre con IHF y Fis en presencia de NarL/NarP en el promotor *nir* de *E. coli*. Pueden también producir remodelaciones en el DNA de un promotor que lo haga más o menos eficiente para la RNAP y de este modo favorecer la activación o la represión del promotor (Browning *et al.*, 2002).

## 2.-TRANSCRIPCIÓN EN EL BACTERIÓFAGO Ø29

Los bacteriófagos, por su simplicidad relativa, son muy útiles como modelo para estudiar múltiples procesos biológicos. De todos los virus que infectan a bacterias Gram-positivas, el fago ø29 es el mejor estudiado.

La familia de bacteriófagos relacionados con ø29 ha sido subclasificada en tres grupos basándose en sus propiedades serológicas, en los mapas genéticos y peptídicos, así como en las secuencias de los extremos 5' del DNA y en las secuencias codificantes de proteínas. El grupo I incluye a los fagos ø29, PZA y ø15, el grupo II a B103, Nf y M2Y, y el tercer grupo tiene un solo componente, GA1. Los fagos pertenecientes a estas familias poseen un genoma y una organización de los promotores muy similar (Pecenкова y Paces, 1999; Yoshikawa, 1986).

El fago ø29 infecta de forma lítica a *B. subtilis*. Es un virus con un genoma de DNA de doble cadena que contiene 19285 pares de bases; posee una proteína terminal unida covalentemente a los extremos 5' del DNA y codifica al menos 20 proteínas.

La expresión de los genes del bacteriófago ø29 ocurre en dos estadios, uno temprano y otro tardío; la transcripción la lleva a cabo la  $\sigma^A$ -RNA polimerasa de *B. subtilis* y es regulada por proteínas tempranas del fago.

Un ciclo infectivo de ø29 comienza con la inyección de su genoma en una célula diana. Los promotores tempranos del virus poseen las secuencias consenso -10 y -35 reconocidas por la subunidad  $\sigma^A$  de la RNAP y, por lo tanto, comienza la transcripción del DNA y la correspondiente síntesis de proteínas. Como resultado se producen las proteínas tempranas, implicadas básicamente en el proceso de replicación del genoma y en la regulación de la transcripción. En un mecanismo mediado por las proteínas tempranas p4 y p6, se activa la transcripción tardía y se produce la represión de la transcripción temprana. Los genes tardíos codifican las proteínas responsables de la morfogénesis de la partícula viral y de la lisis bacteriana. Una vez encapsidados los genomas y lisada la bacteria, se liberan partículas infectivas de ø29 cerrándose, de este modo, el ciclo de infección.

**2.1) Mapa genético y transcripcional de ø29 (figura 6). Cinética de expresión.**

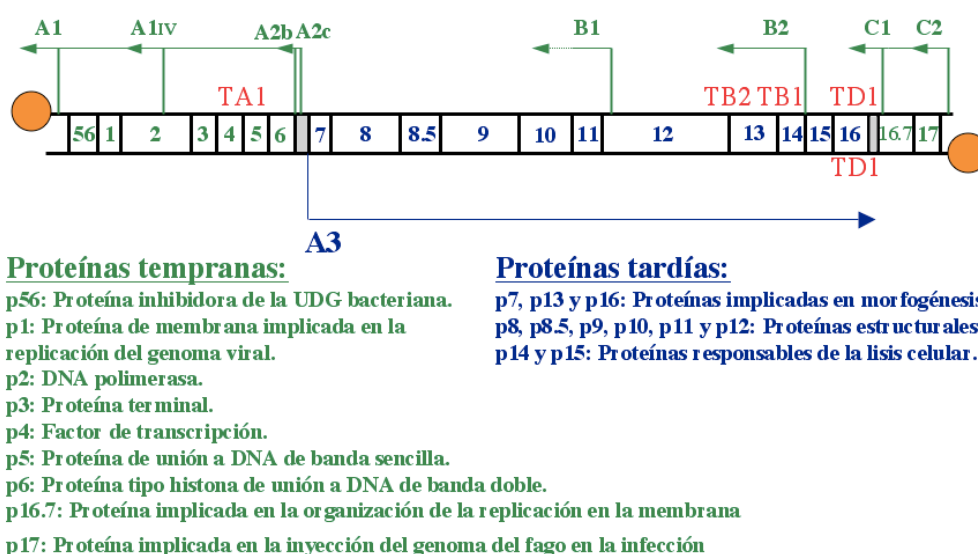


Figura 6: Mapa genético y transcripcional de ø29. Los genes tempranos, así como los promotores desde los cuales se transcriben, están representados en verde. Los genes tardíos y el promotor tardío A3 se muestran en azul. En rojo se señalan los principales terminadores de la transcripción. Los círculos naranjas representan a la proteína terminal unida covalentemente en los extremos 5' del genoma.

Los genes tempranos se encuentran codificados en la misma cadena del DNA y están localizados en ambos extremos del genoma. Los promotores tempranos más importantes son el C2, localizado en el extremo derecho del genoma, y los dos de la zona central, el A2c y el A2b. El promotor C2 es el primero que empieza a transcribir y lo hace antes, incluso, de la entrada completa del DNA en la célula, cuando sólo la mitad derecha del genoma se ha introducido en la bacteria (González-Huici *et al.*, 2006). Desde este promotor tiene lugar la transcripción de los genes 17 y 16.7; existen en esta zona otras fases de lectura abierta que podrían codificar proteínas: 16.9, 16.8, 16.6 y 16.5. Los promotores A2b y A2c son responsables de la transcripción de los genes 6, 5, 4, 3, 2, 1 y 56. Otros promotores también activos son el A1, que dirige la transcripción de un RNA de 174 nucleótidos, llamado pRNA que está implicado en el empaquetamiento del DNA viral en las cápsidas. Los promotores B1 y B2 dan lugar a transcritos antisentido que podrían estar regulando la expresión de determinados genes tardíos (Barthelemy *et al.*, 1988). Desde el promotor A1IV también se transcriben los genes 1 y 56; se trata de un promotor débil, al igual que el C1, que se encuentra localizado a continuación del C2 (Monsalve *et al.*, 1995).

- *Transcripción tardía:*

Los genes tardíos están codificados en la cadena complementaria a la de los tempranos; se transcriben a partir de un solo promotor, el A3, y se encuentran todos agrupados en la región central del genoma. La transcripción transcurre en dirección opuesta a la temprana y da lugar a las proteínas tardías desde p7 a p16, inclusive.

- *Secuencias de los promotores de  $\phi 29$ :*

Los promotores que reconoce el factor  $\sigma^A$  de *B. subtilis* se caracterizan por tener dos secuencias consenso situadas en las posiciones -10 y -35 (TATAAT y TTGACA, respectivamente); aproximadamente el 10% de ellas poseen además el dinucleótido TG en las posiciones -14/-15. La secuencia de los promotores de  $\phi 29$  se muestran en la figura 7.

Promotor	Elemento -35	Espaciador	Elemento -10 extend	Elemento -10	+1
<b>Consenso :</b>	<b>TTGACA</b>	<b>( 17 )</b>	<b>TG</b>	<b>TATAAT</b>	<b>A/G</b>
A1:	TTGACA	( 17 )	TG	TATAAT	A
A1IV:	TTGATA	--	--	--	A
A2c:	TTGAAA	( 17 )	TG	TATAAT	G
A2b:	TTGCAA	( 18 )	TG	TTAAAT	G
A3:	--	--	TG	TATAAT	G
B1:	TTGTTT	( 17 )	TG	TAAACT	A
B2:	TACACA	( 19 )	--	TATAAT	A
C1:	TTTACA	( 17 )	TG	TAAACT	A
C2:	TAGACA	( 18 )	TG	TATACT	G

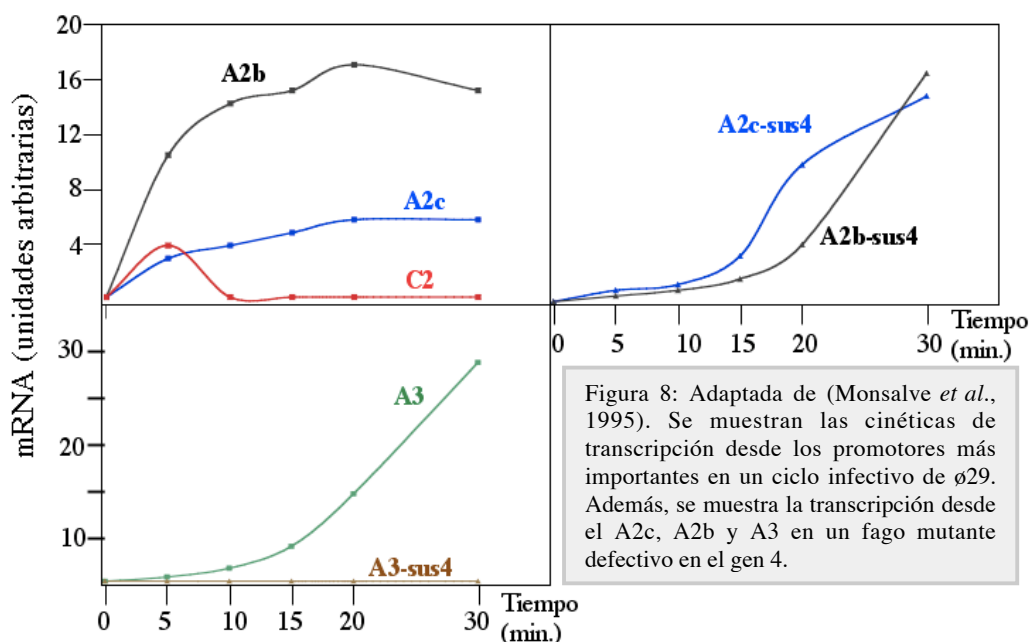
Figura 7: Secuencia de los promotores de  $\phi 29$ , adaptado de (Monsalve *et al.*, 1995). Se señalan los elementos más importantes, incluyendo la base que ocupa la posición +1 del promotor y el número de pares de bases entre los hexámeros -10 y -35. En rojo, se muestran las secuencias consenso.

Todos los sitios de iniciación (+1) son purinas localizadas a 7 u 8 nts del extremo 3' de la región -10; estas son características típicas de los promotores reconocidos por la subunidad  $\sigma^{70}$  de *E. coli* y por su homóloga en *B. subtilis*  $\sigma^A$ . El promotor tardío A3 posee un elemento consenso en -35 muy poco conservado, y aunque posee la secuencia -10 extendida, no es suficiente para que la RNAP de *B. subtilis* se una de forma estable, por lo que se requiere la síntesis de la proteína temprana p4 para su activación. La RNAP de *E. coli*, en cambio, sí reconoce y transcribe el promotor A3 gracias a la presencia de la región -10 extendida. Se analizó

la importancia del dinucleótido  $5'TG^{3'}$  en tres de los promotores de  $\phi 29$  (A1, A2c y A3), y se observó que era importante para el reconocimiento de los tres promotores, para la formación de los complejos cerrado y abierto, pero sobre todo lo era para el promotor A3 que carece de elemento consenso en  $-35$ . No era, en cambio, importante en pasos más avanzados del proceso (Camacho y Salas, 1999).

- *Cinética de transcripción a lo largo del ciclo* (figura 8):

La replicación del genoma resulta evidente a partir del minuto 15 del inicio de la infección, la aparición de nuevas partículas infectivas comienza a partir del minuto 20, aunque no es evidente hasta los 30 minutos, alcanzándose el máximo entre los 40-45 minutos momento en el que, con la lisis celular, termina el ciclo infectivo.



La transcripción desde los promotores A2c y A2b es evidente a los 5 minutos y es máxima aproximadamente a los 15 minutos cuando se alcanza una meseta que se mantiene hasta el final del ciclo. Si los promotores A2c y A2b fuesen totalmente reprimidos cuando está ocurriendo el crecimiento exponencial, la síntesis de proteínas tempranas, que son necesarias para la síntesis del DNA, se vería comprometida. Para evitar esta situación, se mantiene un nivel controlado de la producción de las proteínas tempranas.

Cuando se estudiaron los niveles de transcripción de RNAs de los promotores A2c y A2b utilizando un fago mutante *sus 4*, defectivo en el gen 4, se vio que los niveles de ambos promotores aumentan indefinidamente, comprobándose así la implicación de p4 en su represión (Monsalve *et al.*, 1995).

El promotor A3 es prácticamente inactivo durante los 10 primeros minutos; a partir de los 15 minutos se hace evidente su transcripción, y sigue aumentando hasta el final de la infección. Cuando se analizaron los niveles de transcritos del promotor A3 en un virus mutante *sus 4*, no se detectó ningún tipo de actividad, demostrándose así, *in vivo*, el requerimiento de la proteína p4 para la activación de la transcripción tardía (Monsalve *et al.*, 1995).

La transcripción desde el promotor C2 cae rápidamente a partir del minuto 10 de la infección, ya que es fuertemente reprimido por la proteína temprana p6 (Barthelemy *et al.*, 1989; Monsalve *et al.*, 1995). La proteína p6 es una proteína de unión inespecífica al DNA de doble cadena. Tiene afinidad por los extremos del genoma de

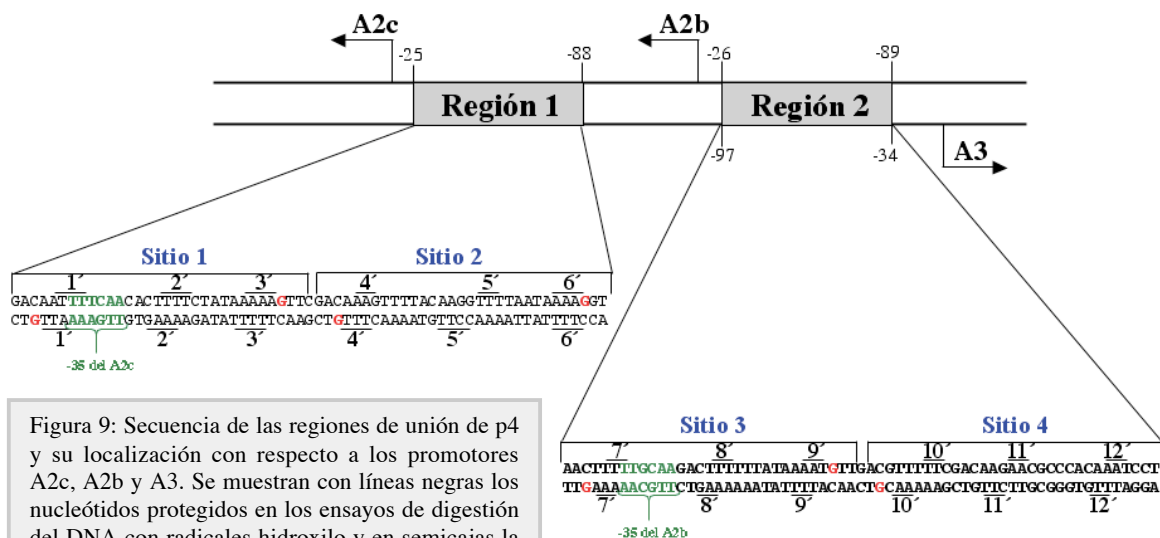
ø29, a los que se une formando un complejo nucleoproteico peculiar (Serrano *et al.*, 1993; Serrano *et al.*, 1990). El mecanismo de represión del promotor C2 mediado por la proteína p6, afecta a la formación del complejo cerrado por competición con la RNAP en la unión al DNA (Camacho y Salas, 2001).

Los promotores de la región central, A2b, A2c y A3, se encuentran agrupados en 219 pb, y la regulación de su actividad está sincronizada. Su funcionalidad depende de la formación de un complejo nucleoproteico en el que están implicadas dos proteínas tempranas del fago, p4 y p6.

## 2.2) Proteína p4:

La proteína p4 es una proteína temprana de unión a DNA. Está codificada por el gen 4, tiene 125 aa y es un dímero en solución. La proteína p4 es un factor de transcripción que, con la colaboración de la proteína p6, regula el paso de la transcripción temprana a la transcripción tardía del fago.

### 2.2.1) Región de los promotores centrales y sitios de unión de p4 (figura 9)



La proteína p4 se une a dos regiones en el genoma de ø29, llamadas región 1 y región 2. Cada una de estas regiones posee 63 pares de bases; la región 1 se encuentra localizada entre los promotores tempranos A2c y A2b, en concreto, desde la posición -25 a la -88 respecto del punto de inicio de la transcripción del A2c; la región 2 está localizada entre el promotor temprano A2b y el tardío A3, y ocupa desde la posición -26 a la posición -89 respecto al punto de inicio de la transcripción del A2b; como se puede observar, la posición de las regiones donde se une la proteína p4 respecto al punto de inicio de la transcripción de los promotores tempranos es muy similar, mientras que la región 2 con respecto al promotor tardío A3, se encuentra una vuelta de hélice más alejada (-34 al -97). En cada región la proteína p4 protege seis grupos de nucleótidos del DNA de la digestión por radicales hidroxilo, llamados 1'-6' en la región 1 y 7'-12' en la región 2 (Camacho y Salas, 2004). A cada grupo de nucleótidos protegidos se le llamó huella, de modo que cada región posee 6 huellas, y la distancia que hay de una huella a otra es de 10 pb en la misma cadena del DNA. Cada una de estas regiones posee dos repeticiones invertidas con una secuencia similar a la consenso, TTG(A/T)AAAA(15)TTTT(T/A)CAA, definida como esencial para la unión de p4



(Barthelemy y Salas, 1989), a las que se llamó sitios. Los sitios 1 y 2 en la región 1 y los sitios 3 y 4 en la región 2.

Cuando se estudió la conformación del DNA en la región 2 se observó la presencia de una curvatura intrínseca de 45° y tras la unión de la proteína p4 se vio que se incrementaba a 85° (Barthelemy y Salas, 1989; Rojo et al., 1990; Serrano et al., 1991).

### 2.2.2) Mecanismo de regulación *in vitro* mediado por la proteína p4

A lo largo de los años, se ha ido profundizando cada vez más en el estudio del mecanismo de regulación de la transcripción del DNA de  $\phi 29$ . El mecanismo por el cual la proteína p4 regula *in vitro* la actividad de los tres promotores de la región central es diferente para cada uno de ellos (figura 10). El sitio 3 de unión de p4 solapa con la región -35 del promotor A2b; la simple unión de la proteína p4 en este sitio afecta a la estabilización de la RNAP en el promotor A2b y por lo tanto, la transcripción temprana desde el mismo se ve disminuída (figura 10a). Esta unión en el sitio 3 es, a su vez, responsable de la activación del promotor A3, debido a que se produce una interacción, mediada por el residuo Arg120, de la p4 con el dominio CTD de la subunidad  $\alpha$  de la RNAP (figura 10b)(Mencía et al., 1996; Mencía et al., 1993). La RNAP de *B. subtilis* no se une de forma estable al promotor A3 porque la homología del elemento -35 es muy baja con respecto al consenso; la interacción de p4 con la subunidad  $\alpha$  hace que la RNAP pueda unirse en este promotor de forma estable, favoreciéndose la formación del complejo cerrado y con ello la transcripción desde el promotor tardío. La unión de p4 al sitio 3 ocurre en la misma cara del DNA que la unión de la RNAP, y este alineamiento estereoespecífico es muy importante para la activación (Serrano et al., 1991). El mecanismo de represión del promotor A2c tiene como mediadora también la interacción entre la p4 (Arg120) y el  $\alpha$ -CTD (figura 10c) (Monsalve et al., 1996). En este caso, la RNAP sí es capaz de reconocer el promotor y es también capaz de formar complejos estables. En presencia de la proteína p4 unida en el sitio 2, se refuerza la interacción de la RNAP en el promotor, impidiendo el escape, aunque sí se produce la síntesis de transcritos abortivos (Monsalve et al., 1998). A pesar de la gran homología entre las subunidades  $\alpha$  de *B. subtilis* y de *E. coli*, la p4 no puede interaccionar con esta última.

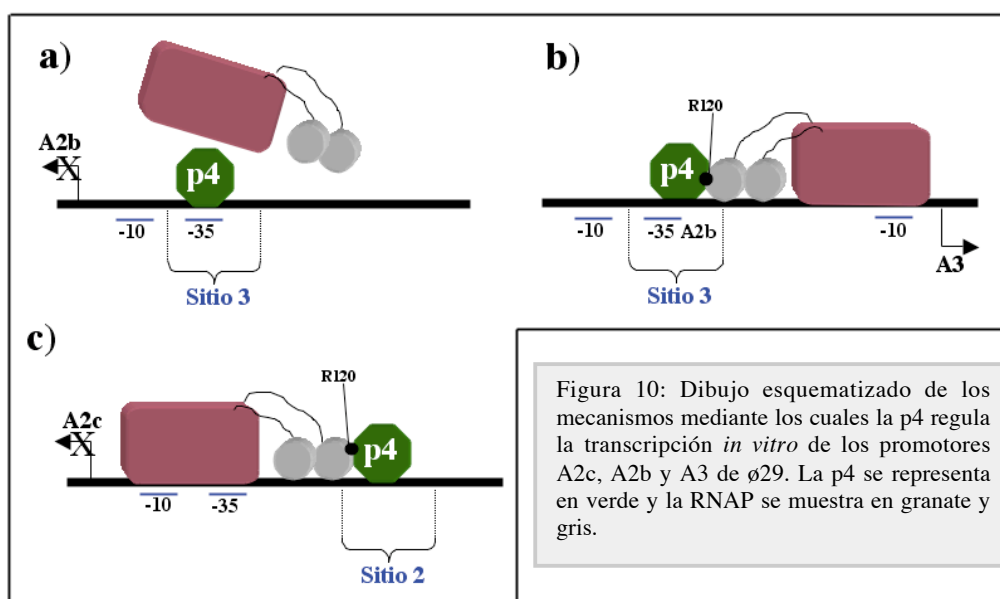


Figura 10: Dibujo esquematizado de los mecanismos mediante los cuales la p4 regula la transcripción *in vitro* de los promotores A2c, A2b y A3 de  $\phi 29$ . La p4 se representa en verde y la RNAP se muestra en granate y gris.



### 2.3) Proteína p6:

La proteína p6 es una proteína viral temprana que interacciona con DNA de doble banda (Freire *et al.*, 1994; Serrano *et al.*, 1990). Es una proteína que recuerda en muchos aspectos a las proteínas nucleoides anteriormente descritas, ya que es pequeña (103 aa), de unión a DNA, se produce abundantemente en las células infectadas por  $\phi 29$  y reconoce, más que una secuencia específica, aspectos estructurales en el DNA. Inicialmente, la unión de p6, que se produce por el surco menor, tiene lugar en regiones del DNA con tendencia a curvarse con una determinada periodicidad. La proteína p6 tiene una especial afinidad por los extremos del genoma de  $\phi 29$ , donde contacta con el DNA cada 12 pares de bases, formando un complejo nucleoproteico muy compacto que es esencial en el inicio del proceso de replicación del DNA viral *in vivo* (Carrascosa *et al.*, 1976; Escarmís *et al.*, 1989). Se ha propuesto que este complejo comprende un núcleo multimérico de p6 que está rodeado por DNA enrollado sobre él (Serrano *et al.*, 1993). *In vitro* p6 es capaz de unirse a lo largo del genoma de  $\phi 29$ , y puesto que en tiempos tardíos de la infección, p6 se encuentra a una concentración 1mM en la célula, hay suficiente cantidad de p6 disponible para recubrir todo el DNA del fago. Todos estos datos sugieren que la p6 tiene un papel estructural en la organización del genoma de  $\phi 29$  (Abril *et al.*, 1997; Gutiérrez *et al.*, 1994).

La proteína p6 está involucrada, además, en procesos de regulación de la transcripción; en primer lugar, es la molécula responsable de la represión del promotor temprano C2 (Barthelemy *et al.*, 1989b) y en segundo, colabora con la proteína p4 en la represión de los promotores tempranos A2c y A2b y en la activación del promotor tardío A3 (Elías-Arnanz y Salas, 1999). Se ha observado que cuando se infectan células con un fago mutante sin sentido en el gen 6, se ve afectada la represión de la transcripción temprana en su totalidad (Camacho y Salas, 2000).

### 2.4) Complejo nucleoproteico p4-p6:

La proteína p4 forma con la proteína p6 un complejo nucleoproteico en la secuencia de DNA donde están localizados los promotores A2c, A2b y A3. La formación de este complejo da lugar a la represión de los promotores tempranos y a la activación del tardío requiriéndose mucha menos proteína p4 para alcanzar mayores niveles de activación y de represión (Elías-Arnanz y Salas, 1999).

Las proteínas p4 y p6 se unen de forma cooperativa al DNA que contiene la secuencia de los promotores. *In vitro*, la proteína p4 se mantiene unida en los sitios 1, 3 y 4 mientras que p6 ocupa la región comprendida entre el sitio 1 y el 3; cuando se aumenta la cantidad de p6 en el ensayo, se observa que el complejo puede extenderse primero por delante del sitio 1 y luego por ambos extremos (Camacho y Salas, 2004) (figura 11a).

En presencia de la RNAP, la p4 es desplazada del sitio 4 para poder llevar a cabo la activación del promotor A3. La p4 se encuentra unida en el sitio 3 y en el sitio 1, y entre ambos se une p6 (Camacho y Salas, 2004) (figura 11b).

En los promotores tempranos, la represión, mediada por el complejo p4-p6, viene dada por el impedimento de la unión de la RNAP. En primer lugar, la subunidad  $\sigma$  no puede interaccionar con los elementos -35 porque coinciden con los sitios 1 y 3 de unión de p4 (-35 del A2c y -35 del A2b, respectivamente), y en segundo lugar, tampoco es posible la interacción de las subunidades  $\alpha$ -CTD con los elementos UP de ambos

promotores, porque su localización coincide con la secuencia recubierta por p6 en el complejo.

Se demostró que la RNAP es capaz de unirse al promotor A2c en presencia del complejo p4-p6, debido probablemente a que el hexámero -10 y el dinucleótido TG se mantienen expuestos, pero no es capaz de llevar adelante la isomerización a complejo abierto (Camacho y Salas, 2004). Por otra parte, se observó que, en ausencia de proteína p6, la represión del promotor temprano A2c depende de la interacción mediada por el residuo Arg120 de la p4 con la RNAP, de modo que, cuando se ensaya la transcripción con la proteína p4 modificada R120Q, este promotor no se reprime. Cuando estos ensayos se llevan a cabo en presencia de las proteínas R120Q y p6, se produce la represión del promotor A2c igual que con la proteína p4 silvestre (Calles *et al.*, 2002). Todos estos datos indican que en la represión mediada por el complejo no influye la interacción de p4 con la RNAP y probablemente esté afectando a la unión de la RNAP en el promotor A2c, siendo tan inestable que se ve impedida la formación del complejo abierto.

En el promotor A3, la proteína p6 estabiliza la unión de la proteína p4 en el sitio 3, lo que permite que también la unión de la RNAP en el promotor sea más estable, resultando más eficiente la isomerización al complejo abierto de transcripción y la consiguiente activación del A3.

La interacción funcional entre la proteína p4 y p6 supone la regulación simultánea de tres promotores distintos, con resultados opuestos (activación/represión).

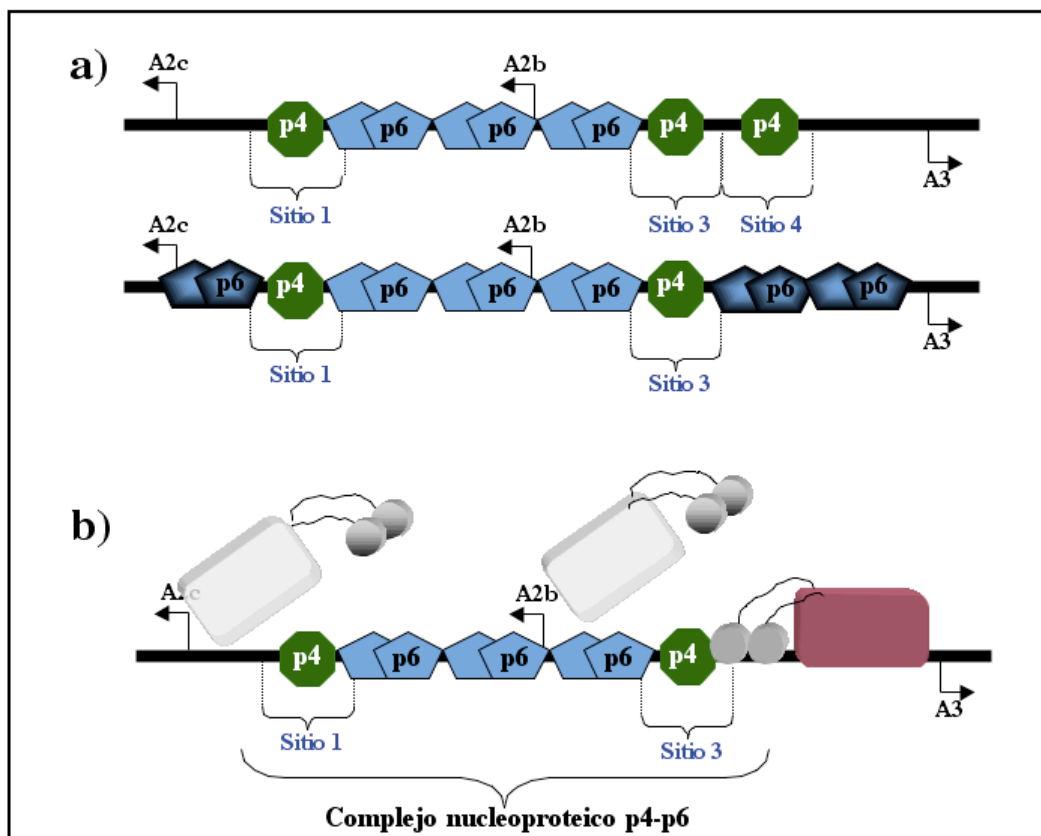


Figura 11: Dibujo esquematizado del complejo nucleoproteico de las proteínas p4 y p6 de ø29. En a) se muestra el complejo *in vitro* en ausencia de RNAP y en b) en presencia de la RNAP. La p4 se representa en color verde y la p6 en azul. La RNAP se muestra en granate y gris, y en azul oscuro se representa el efecto de adicionar dosis altas de p6 en la formación del complejo *in vitro*.

# *OBJETIVOS*

El proyecto propuesto al inicio de este trabajo tenía como fin último profundizar en el mecanismo molecular que regula el paso de la transcripción temprana a la tardía del bacteriófago  $\phi 29$ , así como conocer detalladamente la formación y función de los complejos compuestos por las moléculas, virales y bacterianas, implicadas en el proceso de transcripción y su regulación. Abordamos este proyecto planteando cuatro objetivos principales:

1.- Estudiar la conformación de la secuencia de DNA comprendida entre los promotores A2c y A2b, y de los cambios conformacionales inducidos por la unión de la proteína p4 en esta región.

2.- Definir la secuencia diana de interacción de p4 con el DNA. Caracterizar los determinantes del sitio de unión y de la proteína que confieren especificidad a la interacción p4-DNA.

3.- Determinar la función de cada uno de los sitios de unión de p4 en la regulación del paso de la transcripción temprana a la transcripción tardía, así como, en la formación del complejo regulador p4-p6.

4.- Analizar la generalidad del sistema de regulación de la transcripción, mediante estudio y comparación de la transcripción de Nf, un fago del grupo II de la familia de  $\phi 29$ .

# *MATERIALES Y MÉTODOS*

## 1.- MATERIALES:

### Enzimas y Proteínas:

Las enzimas de restricción, la T4 polinucleótido quinasa, la T4 DNA ligasa, la seroalbúmina bovina (BSA) y la Vent polimerasa, procedieron de New England Biolabs (NEB).

La fosfatasa alcalina, la RQ1 DNasa y el inhibidor de RNasas “Rnasin”, se compraron a Promega, y la polimerasa Pfu turbo a Stratagene.

Las proteínas p4 y p6 de  $\phi 29$  se obtuvieron como se describe en Barthelemy *et al.*, 1987 y Pastrana *et al.*, 1985, respectivamente, y la RNAP de *B. subtilis*, como se describe en Sogo *et al.*, 1979.

### Estirpes Bacterianas:

#### De *E. coli*:

**X11-Blue MRF'** (Sambrook 1989), genotipo:  $\Delta(mcrCB-hsdSMR-mrr)$  173, *endA1*, *supE44*, *thi-*, *recA*, *gyrA96*, *relA1*, *lac*,  $\lambda^-$ , [*F'*, *proAB*, *lacI<sup>q</sup>ZAM15*, *Tn10*, (*tet*<sup>r</sup>)]

**B121(DE3)** (Studier y Moffatt, 1986), genotipo: B *F<sup>-</sup> dcm ompT hsdS(r<sub>B</sub><sup>-</sup>m<sub>B</sub><sup>-</sup>) gal $\lambda$ (DE3)*

**JM 109** (Sambrook 1989), genotipo: *e14<sup>-</sup>(McrA<sup>-</sup>) recA1 endA1 gyrA96 thi-1 hsdR17 (r<sub>k</sub><sup>-</sup>m<sub>k</sub><sup>+</sup>) supE44 relA1  $\Delta(lac-proAB)$  [*F'**traD36 proAB lacI<sup>q</sup>ZAM15*]*

#### De *B. subtilis*:

**110NA** (Moreno *et al.*, 1974), genotipo: (*trp C2- spoA3<sup>-</sup> su<sup>-</sup>*)

### Productos químicos y reactivos:

- La acrilamida (40%) y la bisacrilamida (2%) utilizadas para preparar los geles nativos de poliacrilamida se compraron a Sigma, mientras que la acrilamida 4X de secuenciación que se empleó para los geles desnaturalizantes con urea y la que se utilizó para las electroforesis de proteínas con SDS se compraron a Serva.
- El sephadex G-50 para las columnas de eliminación del nucleótido marcado no incorporado se compró a Sigma.
- La agarosa se compró a Conda y la de bajo punto de fusión a CAMBREX.
- La mayoría de productos como las sales, los ácidos, los alcoholes, la Trizima Base, la urea, el EDTA, la piperidina, etc. se compraron a Sigma o a Merck.
- El kit de extracción de DNA de geles de agarosa *QIAquick* a QIAGEN.
- El kit de purificación de plásmidos *Wizard sv miniprep* a Promega
- Las columnas de fosfocelulosa empleadas para la purificación de proteínas que se utilizaron se compraron a WHATMAN y las de Heparina-Agarosa a Sigma.

### Otros:

- Los antibióticos se compraron a Sigma
- Los nucleótidos y desoxinucleótidos fueron suministrados por PHARMACIA BIOTECH y los marcados radiactivamente con el isótopo <sup>32</sup>P (3000Ci/mmol),  $\alpha$ -dATP,  $\alpha$ -UTP y  $\gamma$ -ATP por AMERSHAM.
- Los oligonucleótidos fueron obtenidos de ISOGEN o Sigma.

## SECUENCIAS DE OLIGONUCLEÓTIDOS:

### A) Oligonucleótidos empleados en clonajes:

pBend-1-Sal I: 5' CACCCTGGTCGACAATTTTC 3'  
 pBend-2-Sal I: 5' TAAATAGAGAACGTCGACAACAACC 3'  
 5GENF-EcoRI: 5' GGTAATGAATTCTTCTCAC 3'  
 3GENF-SalI: 5' CTCCAATGGCGTCGACTCAATCTC 3'  
 Nf1: 5' CATATTTTCCTTCTCTTCC 3'  
 Nf2: 5' CTTACGAAAATCTCGTTCCATCGG 3'

### B) Oligonucleótidos empleados en secuenciación:

SEQ-pBR322: 5' GCGTATCACGAGGCCCTTTCGTCTTC 3'  
 SEQ-PBENDLAU: 5' GCAATTTAACTGTGATAAACTACCGC 3'  
 pT7-C: 5' GGGGCTTTTCACGGTGGACTGC 3'  
 pT7-N: 5' CCGTCTCCGGGAGCTGCATGTG 3'

### C) Oligonucleótidos empleados en amplificación de fragmentos de DNA utilizados como sustratos:

*Oligo 4930*: 5' CCCACTTTCTATTTATATTTGTAG 3'  
*Oligo 28L*: 5' GATTTCTCTCTGCATCATTTTTG 3'  
*Oligo 5218*: 5' TTTCACCTCTTTCTATAATCTATTAGTACTAC 3'  
*Oligo 24XL*: 5' CAAAATATCTTCGTGTTCTTCTGG 3'  
*Oligo A2C 4871*: 5' AACGTTGACGGTTGTCTTTGTG 3'  
*Oligo A3 5063*: 5' TCTACGTTCTCTATTTAACCACACC 3'  
*Oligo 35*: 5' CTGTGTTTGTGTTGATGATGTC 3'  
*OriRin*: 5' GCGCTTTAAAGTAGGGTACAGCGACAAC 3'  
 Nf3: 5' CATATTTTCCTTCTCTTCTTTTCCTTCTCTTTC 3'  
 Nf4: 5' CTCGTTCCATCGGCATATAGCCCATCTCCTTTC 3'  
 Nf 5: 5' GTGTAGTAGGTAACGCCCTTACTACTCTTTTAGTATATCATG 3'  
 Nf 6: 5' CAATAGATTATATTACTATTTTATTATATGACGGAAACCC 3'

### D) Oligonucleótidos empleados en ensayos de interacción con proteínas:

En letras mayúsculas se presenta la parte de la secuencia que pertenece a ø29 y en letras minúsculas, las secuencias inespecíficas. Cada uno de estos oligonucleótidos se hibridó con su complementario. En el caso de los oligonucleótidos mutantes, desde Sit 1A(1) en adelante, solo se presenta la secuencia correspondiente a cada sitio; el resto de la secuencia es igual a la de cada uno de los sitios silvestres correspondientes. Solo se muestra uno de los oligonucleótidos del dúplex; el otro en cada caso será el complementario.

5HA3A: 5' gctcagatctAACTTTTTGCAAGACTTTTTTATAAAATGTTGAC  
 GTTTTTCGACAAGAACGcgctagctcc 3'. (Cinco huellas-región 2)  
 4HA3A: 5' gccagatctAACTTTTTGCAAGACTTTTTTATAAAATGTTGACG  
 TTTTCGGctagctcc 3' (Cuatro huellas-región 2)  
 3HA3A(sitio3): 5' gcactacagatctcAACTTTTTGCAAGACTTTTTTATAA  
 AATGTTGagctagctacatgg 3'  
 3HA3A(sitio4): 5' gcactacagatctcGACGTTTTTCGACAAGAACGCCAC  
 AAATCCTagctagctacatgg 3'

2HA3A:5'ccatgtcgatacagatctAACTTTTTGCAAGACTTTTTTATAta  
tcgctagcactatcgc 3'(Dos huellas-región 2)

3HA2cA(sitio2):5'gcactacagatctcACCTTTTATTAAACCTTGTA  
AACTTTGTCGAgctagctacat 3'

4HA2cA:5'gccagatctAACTTTGTCGAAGCTTTTTATAGAAAAGTGTGAA  
AATTGTCGctagctcc 3' (Cuatro huellas –región 1)

3HA2cA(sitio1):5'gcacagatctTCTCGAAGCTTTTTATAGAAAAGTGTG  
AAATTGTCagctagctacatgg 3'

2HA2cA:5'ccatgtcgatacagatctTTATAGAAAAGTGTGAAAATTGTC  
tatcgctagcactatcgc 3' (Dos huellas-región 1)

Sit 1 A (1): AACTTTTTATAGAAAAGTGTGAAAATTGTT

Sit 1 B (1): AACTTTTTATAAAAAAGTGTGAAAATTGTC

Sit 2 A (1): ACCTTTTATTAAAAAGTGTGAAAACCTTTGTC

Sit 2 B (1): ACCTTTTATTAAAACTTGTAATAATTTGTC

Sit 4 A (1): GACGTTTTTCGACAAGAAATTTATAAAATCCT

Sit 4 B (1): GACGTTTTTCGACAAGAACGCCACAAATGCT

Sitio 3 +1T (1): AACTTTTTGCAAGACTTTTTTTATAAAATGTT

Sitio 3 +2T (1): AACTTTTTGCAAGACTTTTTTTATAAAATGTT

Sitio 3 +3T (1): AACTTTTTGCAAGACTTTTTTTATAAAATGTT

Sitio 3 -1T (1): AACTTTTTGCAAGACTTTTTTATAAAATGTT

Sitio 3 8C (1): AACTTTTGCAAGACTTTTTTTATCAAATGTT

Sitio 3 9C (1): AACTTTGTGCAAGACTTTTTTTATACAATGTTGA

Sitio 3 10C (1): AACTTTGTGCAAGACTTTTTTTATAACATGTTGA

Sitio 3 10T (1): AACTTTATTGCAAGACTTTTTTTATAATATGTTGA

Sitio 3 11C (1): AACTTTGTGCAAGACTTTTTTTATAAACTGTTGA

Sitio 3 11T (1): AACTTTATTGCAAGACTTTTTTTATAAATGTTGA

Sitio 3 12C (1): AACGTTTTGCAAGACTTTTTTTATAAAACGTTGA

Sitio 3 AA : AACTTTTTGCAAGACTTTTTTTATAAAAAGTTGA

Sitio 3 AT : AACATTTTGCAAGACTTTTTTTATAAAATGTTGA

Sitio 3 GA: AACTTTTTGCAAGACTTTTTTTATAAAATATTGA

Sitio 3 GB: AATTTTTTGCAAGACTTTTTTTATAAAATGTTGA

Sitio 3 G: AATTTTTTGCAAGACTTTTTTTATAAAATATTGA

Sitio 3 10 I: AACTTTINOTTTGCAAGACTTTTTTTATAAAATGTTGA

Sitio 3 DAP: AACTTTTTGCAAGACTTTTTTTATAAADAPATGTTGA

#### E) Oligonucleótidos empleados para mutagénesis dirigida:

G11d: 5'TGTTGAGAAATTTCAACACTTTTCTATAAAAAG 3'

G11r: 5'GAAAAGTGTGAAAATTCTCGAACAGGG 3'

G12d: 5'CACTTTTCTATAAAAACCTCGACAAAG 3'

G12r: 5'CCTTGTAAGACTTTGTCGAAGTTTTTATAG 3'

G21d: 5'CACTTTTCTATAAAAAGTTCGAGAAAG 3'

G21r: 5'CCTTGTAAGACTTTCTCGAAGTTTTTATAG 3'

G22d: 5'GGTTTAAATAAAAAGTTGTTGTCTACGTTCTCTATTAAACC 3'

G22r: 5'GAGAACGTAGACAACAACGTTTTATTAAACC 3'

G31d: 5'CCTGTATAAGTTTTTGCAAGACTTTTTTATAAAATGTTG 3'

G31r: 5'CATTTTATAAAAAAGTCTTGCAAAAACCTTATACAGGTGTGG 3'

G32d: 5'CTTTTTTATAAAATCTTGACGTTTTTCGACAAGAACGCC 3'

G32r: 5'GAAAAACGTCAAGATTTTATAAAAAAGTCTTGCAAAAAG 3'

G41d: 5'TTATAAAATGTTGAGGTTTTTCGACAAGAACGCC 3'



G41r: 5'CTTGTCGAAAAACCTCAACATTTTATAAAAAAG 3'

G4+d: 5' CAAGAACGCCACAAATGCTTATGTATCAA 3'

G4+r: 5'CCCTTGATACATAAGCTATTTGTGGCGTTC 3'

## 2.- MÉTODOS:

### 2.1) Clonajes:

#### A) *Plásmido pBend2*

El DNA de la región 1 de unión de p4 se amplificó por PCR utilizando los oligonucleótidos pBend-1-*Sal* I y pBend- 2- *Sal* I y como DNA molde de la reacción, el genoma de ø29. El DNA amplificado de 95 pb se digirió con la enzima *Sal* I y se sometió a electroforesis en gel de agarosa de bajo punto de fusión al 4%, para separar el DNA digerido de 72 pb de tamaño. Se recortó la banda correspondiente al DNA de interés y se purificó utilizando el *kit de extracción de gel QIAquick*. El plásmido pBend2 (Kim *et al.*, 1989) también se digirió con la enzima de restricción *Sal* I y se quitaron los fosfatos 5' terminales utilizando la enzima fosfatasa alcalina. Se separó el plásmido cortado mediante electroforesis en geles de agarosa al 0,9% y se purificó la banda con el DNA digerido y desfosforilado. Se incubaron ambos DNAs en un volumen final de 15 µl, durante 15 h a 16 °C, con la enzima T4 DNA ligasa, siendo la cantidad de plásmido de 100 ng y estableciendo una relación vector:inserto de 1:3 y 1:5. Con el producto de la ligación se transformaron, mediante choque térmico, células de la cepa XL1-Blue de *E. coli*. Las células que habían adquirido el plásmido se seleccionaron por resistencia al antibiótico ampicilina. Se extrajo el DNA plasmídico utilizando el *kit de purificación de DNA Wizard*, y para comprobar que la secuencia era la correcta se sometieron a secuenciación automática en el SIDI (Servicio Interdepartamental de Investigación), utilizando los oligonucleótidos SEQ-pBEND-LAU y SEQ-pBR322. Las células con el plásmido que tenía la secuencia correcta insertada se almacenaron a -70°C. Al plásmido que contenía esta secuencia de 72 pb de la región 1 de unión de p4 de ø29, se le llamó pBend2-wt.

En el plásmido pBend2 se clonaron tres secuencias más de DNA, que contenían diferentes modificaciones en los sitios 1 y 2 de la región 1, Mut 1, Mut 2 y Mut 3. El método empleado fue, básicamente, igual al descrito para la secuencia silvestre, salvo que el DNA molde empleado en la PCR inicial fue el de 3 plásmidos del laboratorio que contenían las secuencias mutadas de la región 1. A los plásmidos obtenidos se les llamó pBend2-Mut1, pBend2-Mut2 y pBend2-Mut3.

#### B) *Proteína p4N*

El gen 4 de Nf se amplificó por PCR utilizando los oligonucleótidos 5GENF-EcoRI y 3GENF-SalI y como molde de la reacción, el DNA genómico de Nf. El producto se digirió con las enzimas *Eco* RI y *Sal* I , al igual que se hizo con el vector de expresión pT7.3. Este vector, una vez digerido, se desfosforiló con la fosfatasa alcalina y se incubó con el DNA de la secuencia del gen 4 de Nf en una relación vector: inserto 1:3 y 1:5 en un volumen final de 15 µl, durante 15 h a 16°C en presencia de la enzima T4 DNA ligasa. Con el producto de la ligación se transformaron por electroporación células JM109 de *E.coli*, y las células que poseían el plásmido se seleccionaron por resistencia a ampicilina. Estas células se crecieron para obtener DNA en grandes cantidades que se utilizó para transformar, por choque térmico, células de expresión de proteínas BL21 de *E. coli*. Se extrajo el DNA plasmídico y se sometió a secuenciación automática en el SIDI utilizando los oligonucleótidos pT7-N y pT7-C. Se vió, tras esta secuenciación, que existían diferencias con la secuencia publicada previamente de este

gen (Mizukami *et al.*, 1986). De modo que se procedió a secuenciar el gen 4 directamente del DNA genómico de Nf disponible en el laboratorio, resultando dichos cambios confirmados; se asumió, por tanto, que la secuencia clonada en el vector de expresión pT7.3 era la correcta.

### C) *Promotores de Nf*

La región central donde se encuentran los promotores A2c, A2b y A3 de Nf, se clonó en un plásmido derivado de pUC18. Es un plásmido que posee a ambos lados de la región de policlonaje secuencias terminadoras de la transcripción y se llama pUC19 $\Omega$ TD1. Se creyó necesaria la utilización de estos terminadores de la transcripción a ambos lados, porque las secuencias de los promotores de Nf son muy similares a las de los promotores de  $\phi$ 29, que son reconocidos y transcritos por la RNAP de *E. coli*; además, dos de los promotores clonados transcriben en un sentido y el tercero, el A3, lo hace en el contrario. El fragmento de interés de 344 pb contiene el DNA comprendido entre la posición +68 del A2c y la posición +60 del A3 y se amplificó por PCR utilizando los oligonucleótidos llamados Nf1 y Nf2, y como molde de la reacción el DNA genómico de Nf. Este DNA se insertó en el sitio *Sma* I del plásmido pUC19 $\Omega$ TD1 y se denominó pUC-NfLP. Con este DNA se transformaron células XL1-Blue de *E. coli* por choque térmico y se seleccionaron las colonias positivas por resistencia a ampicilina. Se extrajo el DNA plasmídico y se secuenció en el SIDI, utilizando los dos mismos oligonucleótidos, Nf1 y Nf2. Las células con la secuencia de interés correcta se almacenaron a  $-80^{\circ}\text{C}$ .

## 2.2) Mutagénesis dirigida:

En el caso de los mutantes de la proteína p4, la mutagénesis se hizo a partir del plásmido pT7.7 donde está clonado el gen 4 de  $\phi$ 29, mientras que cuando las modificaciones se hicieron en la secuencia de DNA de los promotores, se partió del plásmido pUC19 $\Omega$ TD1 que contiene la secuencia comprendida entre la posición +75 del A2c hasta la posición +67 del A3 de  $\phi$ 29.

Se diseñaron para cada mutación dos oligonucleótidos con parte de la secuencia complementaria entre sí y que incluyen el cambio que se va a introducir. En primer lugar se llevaron a cabo en dos tubos separados, por cada mutación, 10 ciclos de PCR; en cada tubo se incluyó uno de los oligonucleótidos (125 ng), el plásmido molde (30 ng), los cuatro dNTPs y la polimerasa termoestable Pfu.

Estos 10 ciclos consistieron en:

- 1) 30 segundos / $94^{\circ}\text{C}$
- 2) 1 minuto a una temperatura igual a la  $T_m$  del oligonucleótido  $-5^{\circ}\text{C}$ , y  $5^{\circ}\text{C}$  menos por cada nucleótido desapareado.
- 3) 8 minutos / $68^{\circ}\text{C}$

Posteriormente se mezcló el contenido de ambos tubos, se adicionó 1  $\mu\text{l}$  de enzima Pfu y se sometió a 20 ciclos más, subiendo la temperatura a la  $T_m$  del oligonucleótido  $-5^{\circ}\text{C}$ .

Tras la reacción de PCR los plásmidos resultantes que fueron íntegramente amplificados, se incubaron con la enzima *Dpn* I, que digiere DNA previamente metilado. *E. coli* metila su DNA tras la replicación para protegerlo de posibles digestiones por enzimas de restricción. El único DNA metilado que debe existir en la reacción es, por tanto, el DNA parental que se ha usado como molde. De este modo, tras este tratamiento solo está presente el plásmido generado mediante PCR que debe contener la modificación producida. Con estos DNAs se transformaron células XL1-Blue de *E. coli*, por choque térmico; los plásmidos con mutaciones en los promotores se

secuenciaron en el SIDI. En el caso de los mutantes de la proteína p4, los plásmidos con la secuencia modificada se introdujeron posteriormente en células de expresión BL21 de *E. coli*, y después se secuenciaron en el SIDI.

### 2.3) Purificación de proteínas:

#### A) *Proteína p4N*

Se creció un cultivo de las células BL21 de *E. coli* que contienen el gen 4 de Nf clonado en el plásmido pT7.3, a 37°C durante una noche. Estas células se utilizaron para inocular medio LB fresco (7,2 litros). Este nuevo cultivo se creció a 37°C con agitación, hasta 60 minutos después de haber alcanzado una densidad óptica de 0,6 unidades medida a 600 nm. Se indujo la expresión de la p4N por adición de IPTG (0,5 mM) y se incubó 30 minutos a 37°C; después se inhibió la transcripción endógena de la bacteria añadiendo rifampicina y se incubó 2 horas más a 37°C. Posteriormente se recogieron las células por centrifugación 15 minutos/2000xg/4°C, obteniéndose 24 g de bacterias.

Se añadió el doble, en peso, de alúmina (48 g) y se machacó en un mortero durante 20 minutos a 4°C. Se resuspendió con 6 volúmenes (144 ml) de tampón B6 (50mM Tris-ClH, pH 7,5, 5% glicerol, 1mM EDTA, 7mM  $\beta$ ME) suplementada con 1M ClNa y se centrifugó 25 minutos/23400xg/4°C, para eliminar la alúmina y las células sin lisar. Al sobrenadante se le añadió 0,3% de polietilenimina (PEI) para precipitar el DNA y se centrifugó 15 minutos/23400xg/4°C. Se añadió sulfato amónico al 70% para precipitar las proteínas y se centrifugó 15 minutos/23400g/4°C; el precipitado se lavó varias veces con B6 con 50% de sulfato amónico y finalmente se resuspendió en B6-0,4M ClNa. La solución se cargó en una columna de fosfocelulosa por caída libre, previamente equilibrada con B6-0,4M ClNa. Se añadieron a la columna varios volúmenes de B6-0,4M ClNa, B6-0,5M ClNa, B6-0,52M ClNa, B6-0,65M ClNa y B6-1M ClNa y se fueron recogiendo las fracciones correspondientes. Las proteínas eluidas se analizaron en geles al 12% de poliacrilamida-SDS (Laemmli, 1970) y se tiñó posteriormente con azul de coomassie. La p4N eluyó en las fracciones de B6-0,5 y 0,52 M ClNa que se juntaron y se diluyeron hasta B6-0,4M ClNa y se pasó por una nueva columna de fosfocelulosa. Se lavó con B6-0,4 y 0,45M ClNa, se recogieron las proteínas eluidas y se corrieron nuevamente en un gel al 12% de poliacrilamida-SDS donde se observó en qué fracciones salía la proteína finalmente pura. Se precipitó con sulfato amónico saturado, se centrifugó 40 minutos/104000xg/4°C y el precipitado se resuspendió en B6-50% glicerol. Se obtuvieron 3,5 ml de p4N, a una concentración de 11,7 mg/ml (con 320 mM de sulfato amónico).

#### B) *Proteína p6N*

La proteína p6 se purificó a partir de la que produce el fago Nf en bacterias infectadas. Las células infectadas fueron *B. subtilis* 110NA. Se preparó un fermentador de 30 litros con medio rico LB y se crecieron células de esta estirpe hasta una densidad óptica de 0,42 unidades medidas a 420 nm y se infectó con el bacteriófago Nf a una multiplicidad de infección de 5. Pasados 30 minutos se centrifugaron las células obteniéndose 40 g. Se añadió el doble en peso de alúmina, se machacó en un mortero 20 minutos a 4°C y se resuspendió en cuatro volúmenes de B6-0,5M ClNa. Se centrifugó 5 minutos/1500xg/ 4°C y 20 minutos/16200xg/ 4°C, para eliminar la alúmina y las bacterias no lisadas. Al sobrenadante se le añadió PEI para eliminar el DNA y se volvió a centrifugar 30 minutos/23400xg/4°C. Las proteínas del sobrenadante se precipitaron con sulfato amónico al 70%, se centrifugó 30 minutos/23400xg/4°C y el precipitado se lavó con B6-40% de sulfato amónico y se resuspendió en B6 a una fuerza iónica de 0,2M. Se pasó la muestra por una columna de fosfocelulosa para eliminar la PEI, ya que

la p6N a alta fuerza iónica no se une a esta columna. Se cargó, a continuación, en una columna de heparina-agarosa a la que se une la p6, y se fueron añadiendo diferentes cantidades de B6-0,2M ClNa, B6-0,25M ClNa, B6-0,28M ClNa, B6-0,3M ClNa y por último B6-0,5M ClNa. Se analizaron las proteínas en un gel en gradiente 10-20% de poliacrilamida-SDS y se juntaron las fracciones donde había más proteína y estaba más pura la p6N, que fueron desde 0,25M a 0,5M ClNa. Se diluyó con B6 hasta 0,19M ClNa y se lavó con B6-0,2M ClNa, B6-0,22M ClNa, eluyéndose con B6-0,5M ClNa. Como quedó algún contaminante, se precipitó con sulfato amónico 70%, y se centrifugó 35 minutos/115000xg/4°C, se resuspendió con B6-50 mM (200 ml) y se bajó un poco más la fuerza iónica, añadiendo 30 ml de glicerol 100%, en estas condiciones p6N se fija a la columna de fosfocelulosa. Se lavó con diferentes volúmenes de B6 y se eligieron las fracciones eluidas entre 100 y 175 mM. Finalmente, se concentró en una columna de heparina-agarosa y se precipitó con sulfato amónico saturado, se centrifugó 35 minutos/115000xg/4°C y se resuspendió en B6-50% glicerol. Se obtuvieron 3 ml de p6N a 15mg/ml (con 700 mM de sulfato amónico).

C) Proteínas modificadas de p4: T4A, Q5A, R6A, H10A, Y33A, K36A, K51A, R54A y K76A.

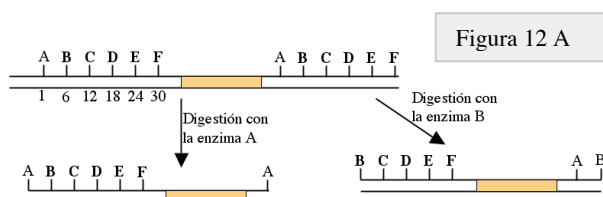
Estas proteínas se purificaron, cada una, partiendo de un cultivo de 3,6 litros; la inducción y sobreexpresión fue muy similar a como se hizo para la p4N y el protocolo de purificación fue como se describe en (Barthelemy *et al.*, 1987).

#### 2.4) Análisis de la existencia de curvatura intrínseca en el DNA de la región 1 de unión de p4:

Se llevó a cabo el análisis en tres fragmentos de DNA de tamaños muy similares; todos ellos se obtuvieron mediante PCR utilizando como molde el DNA genómico de  $\phi 29$ . El fragmento llamado R2 corresponde a 195 pb de la región 2 de unión de p4, que contiene los sitios 3 y 4; los oligonucleótidos utilizados fueron el *oligo 24* y el *oligo 5063*; este fragmento de DNA contiene una curvatura intrínseca de 45° (Barthelemy y Salas, 1989). El fragmento llamado R1 corresponde a un DNA de 200 pb que contiene la región 1 de unión de p4 con los sitios 1 y 2, y se obtuvo utilizando los oligonucleótidos *oligo 4871* y *oligo 5072*. Por último, el fragmento C tiene 198 pb, contiene el promotor C2, carece de curvatura intrínseca, y se obtuvo por PCR con los oligonucleótidos *oligo 35* y *oriRin* y posterior digestión con la enzima *Hinf*I. Estos tres fragmentos se purificaron, se marcaron radiactivamente en sus extremos 5' utilizando [ $\gamma$ -<sup>32</sup>P]ATP y la enzima T4 polinucleótido quinasa que incorpora en el DNA un fosfato en posición 5' y se sometieron a electroforesis en geles nativos al 5% de poliacrilamida en una solución 12 mM Tris-acetato (pH 7,5), 1 mM EDTA. Las electroforesis se corrieron a 8V/cm y a dos temperaturas diferentes, a 65°C y a 4°C; posteriormente los geles se secaron, se expusieron y se revelaron obteniéndose las autorradiografías correspondientes.

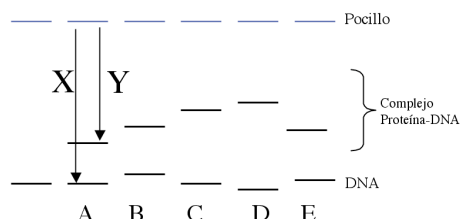
#### 2.5) Permutación circular:

Este tipo de experimentos permiten calcular, de forma aproximada, el centro y grado de curvatura inducido por una proteína en un fragmento de DNA.



Estos ensayos se llevaron a cabo utilizando los plásmidos pBend2-wt, pBend2-Mut1, pBend2-Mut 2 y pBend2-Mut3. Estos

plásmidos tienen flanqueando la región de policlonaaje, donde están insertados los distintos fragmentos de la región 1, secuencias de dianas de restricción colocadas en el



$$\text{Movilidad relativa} = (Y \times 1) / X$$

Figura 12 B

de extracción de gel *QIAquick*. El DNA obtenido se marcó radiactivamente utilizando [ $\gamma$ - $^{32}\text{P}$ ]ATP y la enzima T4 polinucleótido quinasa. Estos fragmentos de DNA se incubaron, durante 15 minutos en hielo, con diferentes concentraciones de proteína p4, en una solución de unión que contenía en un volumen final de 20  $\mu\text{l}$ : 25 mM Tris-ClH (pH7,5), 10 mM  $\text{Cl}_2\text{Mg}$ , 100 mM ClK, 1  $\mu\text{g}$  de poli d[I-C] y 1  $\mu\text{g}$  de BSA. Posteriormente se añadieron 2  $\mu\text{l}$  de glicerol al 30% (v/v) y se cargaron las muestras en geles nativos al 6% de poliacrilamida, corriéndose durante 4 horas a 25 mA y a 4°C. La solución de electroforesis contiene: 12 mM Tris-acetato (pH7,5) y 1 mM EDTA. Los geles se secaron con vacío y calor, se expusieron, y tras revelar las autorradiografías se procedió a hacer las medidas de movilidad relativa de las bandas de DNA. En la figura 12B, se esquematiza el resultado de un experimento de permutación circular. La flecha X representa la movilidad que tiene un determinado fragmento de DNA en ausencia de proteína, la flecha Y representa la movilidad de ese mismo fragmento formando complejo con la proteína. A la movilidad que presenta el fragmento de DNA libre se le asigna el valor 1 y la movilidad relativa es, entonces, el resultado de dividir el valor de Y por el valor de X. De cada uno de estos fragmentos se obtiene un dato de movilidad relativa, que se representa en una gráfica frente al número de nucleótido. En la figura 12C se muestra el resultado tras realizar digestiones hipotéticas. Si se toma como 1 la posición del primer nucleótido tras el corte con la enzima A, entonces, en la gráfica se representará en el eje de ordenadas la movilidad relativa del fragmento digerido con A y en abscisas el valor sería 1; suponiendo 6 nucleótidos entre ambas dianas de restricción tras el corte, la movilidad relativa del fragmento digerido con la enzima B se representaría frente al número de nucleótido 6. Tras la representación, el punto más bajo de la gráfica indica que ese fragmento es el que posee la curvatura más centrada. Conociendo la secuencia de dicho fragmento se puede estimar aproximadamente el par de bases donde está el centro de la curvatura que se genera en el DNA.

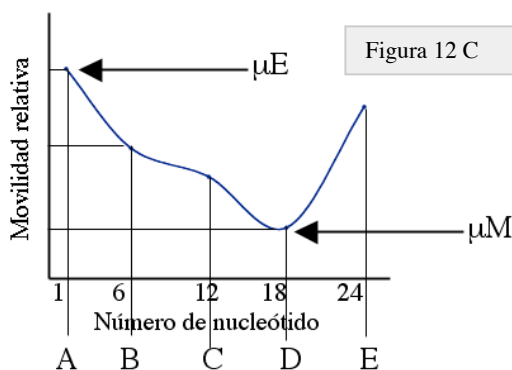


Figura 12 C

$$\mu\text{M (D)} / \mu\text{E (A)} = L \cdot \cos \alpha / 2$$

$\alpha$  = ángulo de curvatura

**$L \cdot \cos \alpha / 2$ .** Donde  $L$  es la longitud de un DNA no curvado y  $\alpha$  es el ángulo de la curvatura que presenta el DNA. La

distancia de lado a lado de un fragmento con la curvatura en el extremo es, virtualmente, la misma que  $L$ . Entonces:

$$\mu M / \mu E = (L \cdot \cos \alpha / 2) / L = \cos \alpha / 2$$

Donde  $\mu M$  es la movilidad del fragmento con la proteína unida en el centro, que es el que posee menor movilidad,  $\mu E$  es la movilidad del complejo con la proteína unida en el extremo y que presenta la mayor movilidad,  $\alpha$  es el ángulo de la curvatura que induce la proteína. En el ejemplo representado en la figura 12C el fragmento  $D$  es el que presenta la menor movilidad relativa ( $\mu M$ ) y el fragmento  $A$  es el que presenta la mayor movilidad ( $\mu E$ ).

En todos los ensayos de permutación circular presentados en este trabajo se muestra el resultado de la autorradiografía y la gráfica correspondiente.

## 2.6) Ensayos de modificación de la movilidad electroforética del DNA:

Estos experimentos se utilizan para estudiar la interacción de proteínas con el DNA y permiten analizar la estabilidad, la afinidad de unión y la formación de complejos del DNA con varias proteínas.

Las reacciones se llevaron a cabo en un volumen final de 20  $\mu$ l de una solución que contiene 25 mM Tris-ClH, pH 7,5 y 10 mM  $Cl_2Mg$ , suplementada con 100 mM de ClK, 1  $\mu$ g de poli[d(I-C)] y 1  $\mu$ g de BSA. El DNA utilizado está marcado radiactivamente y en cada caso se señala su tamaño y método de obtención. Las incubaciones de las proteínas correspondientes con el DNA se hicieron entre 15 y 20 minutos en hielo, se añadieron 2  $\mu$ l de glicerol al 30% y se cargaron en los geles. Se trata de geles nativos al 4 ó 6% de poliacrilamida, en función del tamaño del DNA, que se prepararon en solución de retraso, la misma que se utilizó como tampón de electroforesis y que consiste en Tris-acetato 40 mM, EDTA 1 mM, pH 7,5. Las electroforesis se corrieron a 25 mA de intensidad constante durante 4 horas y a 4°C. Posteriormente, los geles se secaron por vacío y calor y se revelaron mediante autorradiografía.

- Cuando se analizaron por esta técnica fragmentos de DNA que contenían un solo sitio de unión silvestre o modificado y en los ensayos de caracterización del sitio mínimo de unión de p4, los DNAs utilizados fueron oligonucleótidos sintéticos, purificados, marcados e hibridados.

## 2.7) Ensayos de protección a la digestión del DNA por DNasa I:

Para llevar a cabo estos ensayos se requiere que el DNA utilizado esté marcado radiactivamente solo en el extremo 5' de una de las cadenas. Este tipo de fragmentos se obtiene mediante PCR, en la cual uno de los oligonucleótidos se marca con  $[\gamma\text{-}^{32}P]ATP$ . El producto amplificado se purifica mediante electroforesis en geles de agarosa y posterior extracción del DNA con el *kit de extracción de gel QIAquick*. En cada ensayo realizado de este tipo se indica el tamaño del DNA amplificado y el extremo que se marcó.

Las reacciones se llevaron a cabo en un volumen final de 20  $\mu$ l, con aproximadamente 15000 cpm de DNA marcado, en un tampón con 25 mM Tris-ClH, pH 7,5, 10 mM  $Cl_2Mg$ , suplementado con 1  $\mu$ g de poli[d(I-C)] y con 50 mM de ClK en los ensayos con DNA de Nf y 100 mM en los ensayos con DNA de  $\phi 29$ ; además, se adicionó 1  $\mu$ g de BSA, excepto en aquellos casos en los que se incluían las proteínas p6 o p6N. Las proteínas con el DNA se incubaron durante 20 minutos a 37°C, y se inició la digestión añadiendo 0,05 unidades de la enzima RQ1 DNasa. Tras 2 minutos de incubación a 37°C, la reacción se paró con EDTA (20 mM final) y el DNA se precipitó

con 0,3 M de acetato potásico, 10 µg de tRNA y 5 volúmenes de etanol 100%. Tras lavar el precipitado, se resuspendió con 8 µl de solución de formamida (95% formamida, 20 mM EDTA). Las muestras se hirvieron durante 3 minutos y se cargaron en geles desnaturizantes al 6% de poliacrilamida-8M urea. La electroforesis se llevó a cabo a una potencia de 70 vatios constantes y durante un tiempo que osciló entre 1 y 3 horas en función de la región de interés en cada DNA ensayado. Tras secarlos y exponerlos, se revelaron por autorradiografía.

- Existe un ensayo de competición en el DNA de Nf de la unión de la RNAP y los complejos p4N-p6N y p4-p6, que se desarrolló básicamente como los ensayos anteriormente descritos con la única diferencia de que se permitía la formación del complejo p4N-p6N ó p4-p6 con el DNA durante 10 minutos y posteriormente se adicionaban concentraciones crecientes de la RNAP, se incubaban 15 minutos a 37°C y se añadía la RQ1 DNasa; a partir de este paso el ensayo se llevó a cabo como se ha descrito antes.

## **2.8) Secuenciación del DNA (Maxam y Gilbert, 1980)**

Estos ensayos se llevaron en paralelo a cada experimento de protección del DNA a la digestión con DNasa I, para poder caracterizar las posiciones, en la secuencia, de las bandas protegidas o digeridas. El DNA previamente marcado en uno de sus extremos (100000 cpm) se incubó con ácido fórmico durante 7 minutos a temperatura ambiente, se precipitó el DNA con acetato sódico 0,3 M y 5 volúmenes de etanol 100%, se incubó 5 minutos en un baño con nieve carbónica y etanol, se lavó el precipitado, se secó con vacío y se resuspendió en 30 µl de piperidina 1M. Posteriormente, se incubó 30 minutos a 65°C, se secó con vacío y el precipitado se lavó varias veces con agua, se resuspendió con solución de formamida, y se cargó en geles desnaturizantes al 6% de poliacrilamida-8M urea con las muestras de cada ensayo de protección del DNA a la digestión con DNasa I.

## **2.9) Ensayos de transcripción truncada *in vitro***

Las reacciones en estos ensayos se llevaron a cabo en un volumen final de 25 µl, que contenían 25 mM Tris-ClH, pH7,5, 10 mM Cl<sub>2</sub>Mg, 2 mM DTT, 7,5 unidades de RNasin, 1,25 µg de poli[d(I-C)], y 1,25 µg de BSA salvo en las muestras con p6 y p6N. Se suplementó además con 100 mM de ClK en los ensayos con el DNA de ø29 y con 50 mM de ClK cuando se trataba del DNA de Nf. Se añadieron 2 ng de los fragmentos de DNA molde, indicados en cada experimento, que contenían los tres promotores, A2c, A2b y A3 de ø29 o de Nf. Se incubó con la RNAP de *B. subtilis* y las proteínas señaladas en cada caso durante 15-20 minutos a 37°C. Se inició la reacción de transcripción con la adición de 200 µM de CTP, GTP, ATP, 50 µM de UTP y 1µCi de UTP (α -<sup>32</sup>P). Se incubó 20 minutos a 37°C y se paró la reacción con la adición de 30 µl de EDTA 50 mM. Los nucleótidos no incorporados en la reacción se eliminaron por medio de una columna de 1 ml de Sephadex G-50. Tras secar las muestras con vacío, se resuspendieron en 10 µl de solución de formamida, se hirvieron 3 minutos y se cargaron en un gel desnaturizante al 6% de poliacrilamida-8M urea. La electroforesis se realizó a temperatura ambiente a una potencia de 70 vatios constantes durante 1 hora; los geles se secaron, se expusieron y se revelaron por autorradiografía.

# ***RESULTADOS***

---



## 1.- DETERMINACIÓN DEL SITIO MÍNIMO DE UNIÓN DE p4

Como se ha descrito en la introducción de este trabajo, la proteína p4 se une a dos regiones en el DNA de  $\phi 29$  (figura 13A). Cada una de estas regiones posee dos secuencias con repeticiones invertidas en sus extremos, según la secuencia consenso TXGAAAAA(15 pb)TTTACXA (Barthelemy y Salas, 1989). En estas regiones del DNA se generan *huellas* debido a la protección de nucleótidos por p4 frente a la ruptura química por radicales hidroxilo (Camacho y Salas, 2004). Se le designó un número *prima* a cada una de estas huellas en función de su posición en el DNA respecto al promotor A2c; así, la región 1 incluía las huellas 1' - 6' y la región 2 las huellas 7' - 12' (figura 13A).

Teniendo en cuenta que las huellas eran producidas por la unión de la proteína p4, nos propusimos analizar cuál era la secuencia mínima de DNA que la proteína requería para unirse y así poder determinar cuál era su sitio de unión. Se diseñaron fragmentos de DNA que contenían todas las huellas de cada región y fragmentos que carecían de una huella [(1' - 5') y (7' - 11')], dos huellas [(1' - 4') y (7' - 10')] y así sucesivamente. Salvo en aquellas moléculas que contenían las 6 huellas de cada región, la secuencia de  $\phi 29$  eliminada en los fragmentos era sustituida por secuencia inespecífica del mismo tamaño. Estos fragmentos de DNA se marcaron radiactivamente y se incubaron con la proteína p4; el resultado de la formación de complejos se puede ver en la figura 13 B y C.

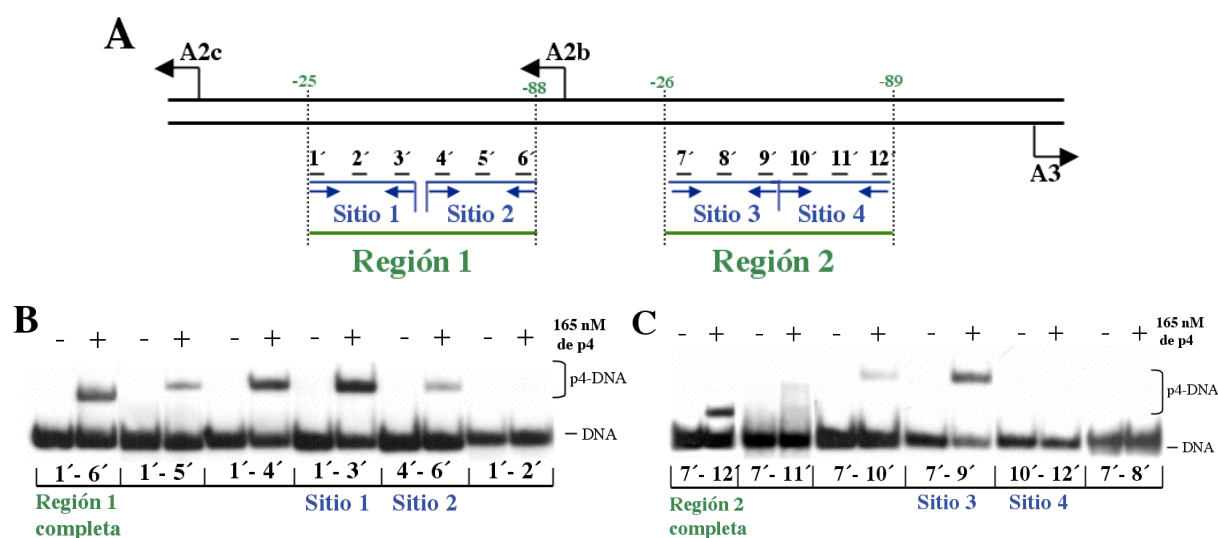


Figura 13: En A se muestra un esquema de la organización de la región de unión de la proteína p4 en el genoma de  $\phi 29$ . El sentido de la transcripción temprana y tardía se representa por medio de flechas negras. Las líneas verdes representan la secuencia de las regiones 1 y 2; su posición en relación a los promotores A2c y A2b, respectivamente, se muestra con números verdes sobre el esquema. Con líneas y flechas azules se simbolizan las repeticiones invertidas en los extremos de la secuencia. En líneas y números negros del 1' al 12' se muestran las *huellas* producidas por la protección del DNA a la ruptura por radicales hidroxilo. En B y C se presenta el resultado de los ensayos de determinación de la secuencia mínima requerida para la unión de p4 en la región 1 (B) y en la región 2 (C). Se trata de fragmentos de DNA que contienen la secuencia de una región completa (1' - 6') y (7' - 12') o la secuencia que resulta de ir eliminando una *huella* progresivamente. Estos fragmentos marcados radiactivamente se incubaron con una concentración de 165 nM de p4, y la formación de los complejos se resolvió en geles nativos al 6% de poliacrilamida a 4°C.

La proteína p4 reconoce y se une a cada una de las regiones completas de forma independiente [(1' - 6') y (7' - 12')].

En el caso de la región 1, todos los fragmentos analizados, salvo aquél que solo contenía 2 huellas (1' - 2'), interaccionaron con la proteína p4 dando lugar a la aparición de una banda de DNA de menor movilidad, correspondiente al complejo p4-DNA. La proteína p4 requiere, al menos, las tres huellas consecutivas, que incluyen una repetición invertida. La región 1 contiene dos sitios de unión de p4 independientes, los llamados sitio 1 y sitio 2 con 3 huellas y una repetición invertida cada uno (figura 13B).

En la región 2, la proteína p4 sólo es capaz de unirse al DNA cuando en el fragmento están incluidas las tres huellas, 7' - 9', que forman el sitio 3. Además, sorprendentemente, la cantidad de DNA retrasado disminuye al aumentar la cantidad de secuencia de DNA de 29, como se puede ver en la figura 13C si se compara el carril que contiene la repetición invertida del sitio 3 (7'-9') con los carriles que contienen 4 (7' - 10') y 5 (7' - 11') huellas. Teniendo en cuenta la existencia de una curvatura intrínseca en esta región (Barthelemy *et al.*, 1989), es posible que en estos fragmentos se produzcan modificaciones conformacionales del DNA que interfieran en la unión de la proteína, en la estabilidad del complejo, o en el reconocimiento de la secuencia por p4. No existe unión detectable de p4 al sitio 4 (10' - 12'); la unión de p4 a este sitio ha sido demostrada mediante ensayos de protección del DNA a la digestión con DNasa I y con radicales hidroxilo, en fragmentos de DNA que contenían toda la región 2, que incluían, por tanto, los dos sitios, el 3 y el 4 (Camacho y Salas, 2004). Por tanto, la unión al sitio 4, o bien depende de la previa unión al sitio 3, o es inestable y el complejo no resiste la fuerza a la que se somete durante la electroforesis, por lo que no sería detectable por esta técnica.

De estos resultados podemos concluir que el sitio mínimo de unión de la proteína p4 se caracteriza por contener en su secuencia una repetición invertida que incluye 3 huellas de protección a la digestión por radicales hidroxilo.

## 2.- AFINIDAD RELATIVA DE p4 POR SUS SITIOS DE UNIÓN

La proteína p4 tiene cuatro sitios de unión en el DNA de  $\phi 29$ . Puesto que resultados anteriores sugerían que la proteína no se unía con la misma afinidad a cada uno de ellos, se decidió estudiar la formación del complejo p4-DNA en cada uno de los sitios. Se utilizaron cantidades iguales de fragmentos de DNA de 60 pb, que contenían secuencias de 31 pb correspondientes a cada sitio y en ambos extremos secuencias de DNA inespecíficas pero iguales para todos ellos. Se incubó cada fragmento de DNA con las cantidades de proteína p4 indicadas en la figura 14.

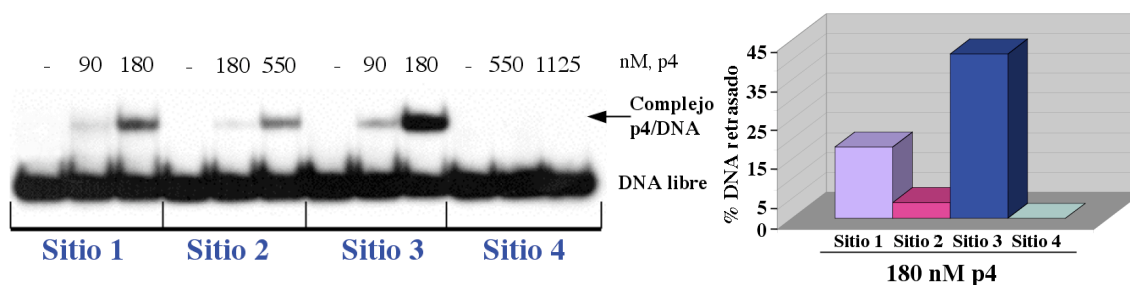


Figura 14: Estudio de la afinidad de p4 por cada uno de sus sitios. La autorradiografía muestra el resultado de la formación de complejos p4-DNA con una concentración fija (4 nM) de cuatro fragmentos distintos de DNA y concentraciones crecientes de la proteína p4. Estos fragmentos tienen el mismo tamaño, están marcados radiactivamente y contienen 31 pb de la secuencia de cada uno de los sitios de unión de p4. El porcentaje de DNA formando complejo se cuantificó y el resultado obtenido para la concentración de 180 nM de p4 se representa en la gráfica.

La concentración de proteína requerida para la formación del complejo p4-DNA en los sitios 1, 2, 3 y 4, y la proporción de complejo formado, varía dependiendo del sitio de unión analizado. La cantidad de DNA formando complejo fue cuantificada y expresada en términos de porcentaje de DNA retrasado. El resultado mostró que la formación de complejos p4-“sitio de unión”, es prácticamente el doble con la secuencia del sitio 3 que con la del sitio 1 (para 180 nM, 41,5 % y 18%, respectivamente) y aproximadamente cinco veces mayor con el sitio 1 que con el sitio 2 (18% y 4% para 180 nM de p4), no obteniéndose formación del complejo en el sitio 4 con cantidades altas de p4 (figura 14).

### 3.- CONFORMACIÓN DEL DNA DE LA REGIÓN 1

#### 3.1) Curvatura intrínseca:

En estudios previos al desarrollo de este trabajo se determinó que el DNA de la región 2 localizado entre los promotores A2b y A3 poseía una curvatura intrínseca (Barthelemy *et al.*, 1989). Posteriormente se analizó cómo era la conformación de esta región tras la unión de la proteína p4. Se determinó que, siendo la curvatura intrínseca de aproximadamente unos 45°, la unión de la proteína p4 acentuaba esta curvatura hasta alcanzar 85° (Rojo *et al.*, 1990). El centro de la curvatura inducida, que coincide con el centro de la curvatura intrínseca, está localizado alrededor de la posición -58 respecto al promotor A3, es decir, en el sitio 4 en su parte más próxima al sitio 3. Estos datos hacían pensar que la unión de p4 al DNA podría requerir un DNA intrínsecamente curvado, de modo que fuese necesaria la presencia de una curvatura para la unión de la proteína a sus sitios. Puesto que la proteína p4 tiene otros 2 sitios de unión en la región 1, nos propusimos estudiar la conformación del DNA en esta región. Para ello, elegimos un fragmento de DNA de 200 pb que contenía el sitio 1 y el sitio 2, es decir, la región 1 completa (R1). Se incluyó también en el ensayo un fragmento de 195 pb con la región 2 que contiene los sitios 3 y 4 (R2) y que posee la curvatura intrínseca de 45° previamente definida, sirviendo por tanto, como control positivo, y por último se utilizó, como control negativo, un fragmento con la secuencia del promotor C2, de 198 pb, y del que se sabía que carecía de curvatura intrínseca (C).

Cuando en una secuencia de DNA existe una curvatura intrínseca, su conformación resulta más rígida a bajas temperaturas que la de un fragmento de DNA con una secuencia no curvada. Esto se pone de manifiesto cuando dichos fragmentos de DNA se someten a electroforesis en geles nativos de poliacrilamida a bajas temperaturas, en comparación con lo que ocurre con estos mismos fragmentos en una electroforesis idéntica pero a altas temperaturas (Crothers y Drak, 1992; Dieckman, 1989; Kahn, 1999). En una electroforesis a 65°C, cada fragmento de DNA se mueve en función de su tamaño, mientras que a 4°C la movilidad depende de la existencia de una determinada conformación en la secuencia.

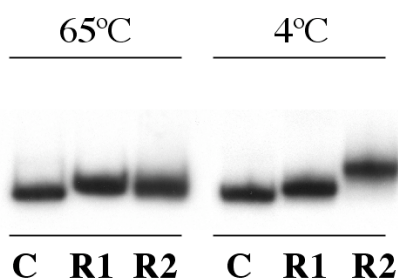


Figura 15: Ensayo de movilidad electroforética dependiente de temperatura para el estudio de la curvatura intrínseca. Tres fragmentos de DNA obtenidos por PCR y marcados radiactivamente se sometieron a electroforesis en geles nativos al 5% de poliacrilamida a dos temperaturas diferentes, 4 y 65°C.

Como se puede ver en la figura 15, cada uno de estos fragmentos de DNA migra a 65°C en función de su tamaño. En cambio, a 4°C los fragmentos C y R1 mantienen la misma movilidad, mientras que el fragmento R2 presenta una movilidad anormalmente reducida.

Los resultados de este ensayo nos permiten deducir dos cosas; en primer lugar que la región 1 carece de una curvatura intrínseca significativa, y en segundo lugar que

la presencia de una curvatura previa en el DNA no es un requerimiento para la unión de la proteína, puesto que la p4 se une a la región 1 y ésta carece de curvatura intrínseca.

### 3.2) Curvatura inducida:

La curvatura final que se genera tras la unión de la proteína p4 a la región 2 es de 85° (Rojo *et al.*, 1990). Esta curvatura se ha interpretado como una acentuación de una curvatura intrínseca ya existente en el DNA. Aunque la región 1 carece de curvatura intrínseca, nos propusimos averiguar si la proteína p4, al unirse a los sitios 1 y 2, inducía también una curvatura en el DNA. Para ello, llevamos a cabo ensayos de permutación circular. La permutación circular se basa en que la movilidad, en geles de poliacrilamida nativos a 4°C, de un fragmento de DNA curvado, varía en función de la posición que el centro de curvatura tiene respecto a los extremos del DNA; de modo que, a igualdad de tamaño, un fragmento con la curvatura en el extremo del DNA, tiene una mayor movilidad que uno con la curvatura en el centro de la secuencia.

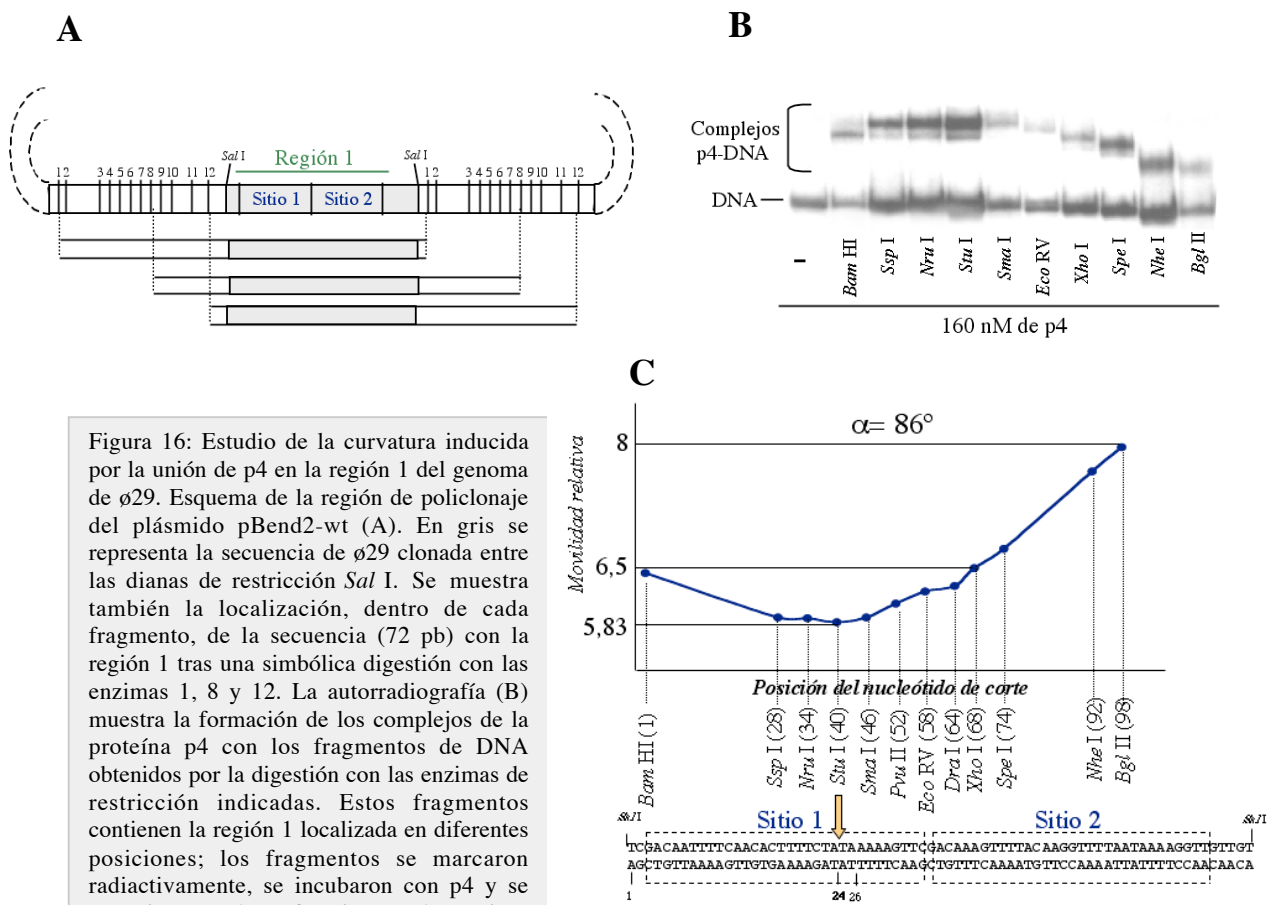
Se insertó en el sitio de restricción *Sal* I del plásmido pBend2 (Kim *et al.*, 1989) una secuencia de 72 pb con la región 1 del DNA de ø29. Al plásmido obtenido se le llamó pBend2-wt y contiene los sitios 1 y 2 de unión de p4 en la orientación indicada en la figura 16A. Este inserto de DNA se encuentra flanqueado por secuencias de dianas de enzimas de restricción iguales a ambos lados y colocadas en el mismo orden. La digestión con las enzimas puede dar lugar a fragmentos de dos tamaños, de 191 pb cuando la digestión genera extremos romos, o de 187 pb cuando se trata de extremos cohesivos. Al cortar con las diferentes enzimas se obtienen fragmentos de tamaños muy similares que difieren en la posición del fragmento de 72 pb respecto a los extremos del DNA (figura 16A).

Se obtuvieron 12 fragmentos diferentes con la secuencia de 72 pb permutada. Estos DNAs purificados y marcados radiactivamente se incubaron con la proteína p4 y los complejos formados se resolvieron mediante electroforesis a baja temperatura en geles nativos de poliacrilamida. La movilidad que presentan todos los fragmentos de DNA en ausencia de proteína es muy similar (figura 16B), lo que confirma la ausencia de una curvatura intrínseca significativa en esta región. En cambio, como se puede ver en la figura 16B, se obtiene un gradiente de movilidades en los fragmentos de DNA que están formando complejo con la proteína, lo cual significa que existe una curvatura en el DNA inducida por la unión de p4. El fragmento obtenido tras la digestión con *Stu* I es el que presenta menor movilidad, seguido por el obtenido con la digestión de *Nru* I, mientras que el fragmento digerido con *Bgl* II es el de mayor movilidad.

Para determinar la localización del centro de la curvatura inducida en esta región, se representaron en una gráfica los valores de la *movilidad relativa* de cada uno de estos fragmentos, frente a la *posición del nucleótido* que marca el punto de corte de cada enzima de restricción (ver Materiales y Métodos). La *movilidad relativa* es el resultado de la razón entre la distancia recorrida por cada fragmento formando el complejo con la proteína y la distancia que recorre la molécula de DNA libre. La *posición de los nucleótidos* se determinó tomando como 1 la posición del primer nucleótido que se obtiene tras digerir el fragmento con la enzima *Bam* HI y, a continuación, numerando cada nucleótido (de la secuencia del plásmido) hasta el punto de corte de la siguiente enzima, *Ssp* I (28), y así sucesivamente (figura 16C).

Una vez representados los datos, el centro de curvatura viene dado por la posición central del fragmento que marque el mínimo en la curva de la gráfica (figura 16C). Como el fragmento de menor movilidad es el obtenido con la digestión de *Stu* I, el centro de la curvatura, por extrapolación en la gráfica, estaría entre los pares de bases 25/26, contando desde la izquierda de la secuencia clonada; puesto que *Nru* I es el

siguiente con menor movilidad relativa y estimando un error límite de  $\pm 2$  pb, consideramos como centro aproximado de la curvatura la posición 24.



El centro de curvatura se encuentra localizado en el sitio 1, pero no en el centro del sitio, sino que está ligeramente desplazado hacia el sitio 2 (figura 16C). Algo muy similar se describió en la región 2, donde el centro de curvatura está en el sitio 4 pero desplazado hacia el sitio 3 (Rojo *et al.*, 1990).

Para determinar el ángulo de la curvatura inducida por p4 en la región 1, se usó el método de Thompson y Landy (Thompson y Landy, 1988), que implica introducir en una fórmula matemática (ver Materiales y Métodos) los valores de las medidas de movilidad relativa de los complejos de mayor y menor migración. En el caso de la región 1, el complejo de mayor movilidad y que, por tanto, presenta la curvatura más próxima al extremo del DNA es el obtenido por la digestión con *Bgl* II, mientras que el complejo de menor movilidad y que, por tanto, posee la curvatura más centrada, es el fragmento obtenido por la digestión con *Stu* I. El resultado, tras aplicar la fórmula matemática, reveló que el grado de curvatura era de  $86^\circ$ , que es prácticamente el mismo que presenta el DNA en la región 2 ( $85^\circ$ ) tras la unión de la proteína p4 (Rojo *et al.*, 1990). Es interesante que en la región 1, que carece de curvatura intrínseca, p4 induzca el mismo grado de curvatura que en la región 2, donde el DNA está curvado  $45^\circ$ . Estos

resultados sugieren que la proteína p4 debe inducir una curvatura de aproximadamente 85° para poder unirse de forma estable a un DNA que posee dos de sus sitios de unión.

### 3.3) Influencia de la secuencia de cada sitio en la inducción de la curvatura por p4 en la región 1:

Con la intención de estudiar cuáles eran los requerimientos de la proteína para unirse al DNA y cuál era la implicación de cada uno de los sitios en la inducción de la curvatura, analizamos la generación de la misma en tres derivados de DNA:

	Sitio 1	Sitio 2
<b>Silvestre</b>	TCGACAATTTTCAACACTTTTCTATAAAAAGTTCTGACAAAGTTTACAAGGTTTAAATAAAAGGTTGTTGT AGCTGTTAAAAGTTGTGAAAAGATATTTTCAAGCTGTTTCAAAATGTTCCAAAATTTATTTCCAAACAACA	
<b>Mut 1</b>	TCGACAATTTTCAACACTTTTCTATAA <b>TAAC</b> TTCTGACAAAGTTTACAAGGTTTAAATAAAAGGTTGTTGT AGCTGTTAAAAGTTGTGAAAAGATATTT <b>TAAG</b> CTGTTTCAAAATGTTCCAAAATTTATTTCCAAACAACA	
<b>Mut 2</b>	TCGACAATTTTCAACACTTTTCTATAAAAAGTTCTGACAAAGTTTACAAGGTTTAAATA <b>CAATT</b> TGTTGT AGCTGTTAAAAGTTGTGAAAAGATATTTTCAAGCTGTTTCAAAATGTTCCAAAATTTAT <b>GTAAA</b> ACAACA	
<b>Mut 3</b>	TCGACAATTTTCAACACTTTTCTATAAAAAG <b>CT</b> CTG <b>AGAA</b> AGTTTACAAGGTTTAAATAAAAGGTTGTTGT AGCTGTTAAAAGTTGTGAAAAGATATTTT <b>CGAG</b> CTCTTTTCAAAATGTTCCAAAATTTATTTCCAAACAACA	

Figura 17: Secuencias de los fragmentos silvestre y mutados que se clonaron en el plásmido pBend2. En rojo se muestran los nucleótidos modificados en cada uno de ellos.

Un derivado con mutaciones en el sitio 1 que se denominó **Mut 1**, otro con mutaciones en el sitio 2, **Mut 2**, y un tercer derivado con mutaciones en ambos sitios, **Mut 3** (figura 17). **Mut 1** posee dos pares de bases modificadas dentro del sitio 1, involucradas en la interacción p4-DNA, como se ha podido ver en ensayos de interferencia por metilación (Monsalve *et al.*, 1998) y mediante ensayos de protección a la digestión por radicales hidroxilo (Camacho y Salas, 2004). **Mut 2** posee tres nucleótidos cambiados dentro del sitio 2 (figura 17); dos de estas mutaciones también se ha visto que están involucradas en la unión de la proteína al DNA. Y por último, **Mut 3** tiene dos pares de bases cambiadas; se trata de una timina por una citosina en el sitio 1, cuya implicación en la unión de p4 al DNA no ha sido determinada, y una citosina por una guanina en el sitio 2 implicada en la unión al DNA (Monsalve *et al.*, 1998).

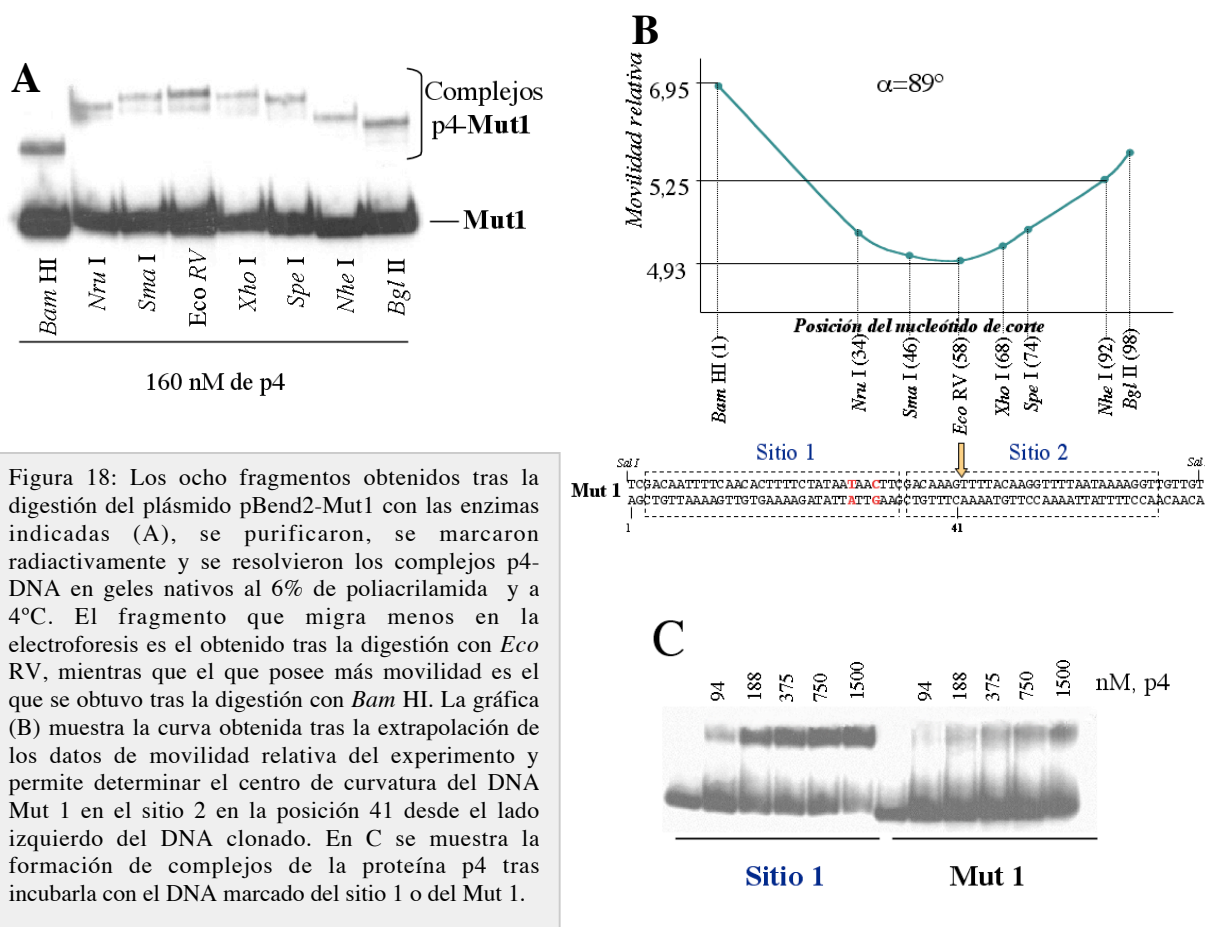
Para estudiar la curvatura inducida por la unión de p4 en la región 1 cuando existen modificaciones en sus sitios de unión, se clonaron los fragmentos mutados Mut 1, Mut 2 y Mut 3 en el plásmido pBend-2 (p-Bend2-Mut1, pBend2-Mut2 y pBend2-Mut3) y se llevaron a cabo ensayos de permutación circular para analizar el centro y grado de curvatura generados por p4.

Cuando se analizó el DNA **Mut 1** (figura 18A), se observó que el fragmento digerido con *Bam* HI era el de mayor movilidad, mientras que el fragmento digerido con *Eco* RV fue el que mostró menor movilidad. Teniendo en cuenta que la orientación en la que se encuentra Mut1 en el plásmido es la misma que la del DNA salvaje, esta diferencia en las movilidades de los fragmentos de restricción implica un cambio en el centro de curvatura.

Tras extrapolar los datos obtenidos de la movilidad relativa de los fragmentos a la gráfica (figura 18B), se observó que el centro de curvatura estaba aproximadamente en el par de bases 41, contando desde la izquierda del fragmento, estando localizado, por tanto, en el sitio 2. El grado de curvatura aparente generado por la proteína p4 unida



a este DNA mutado en el sitio 1 fue de  $89^\circ$ . El resultado indica que al mutar el sitio 1 se mantiene aproximadamente el mismo grado de curvatura en el DNA pero se modifica el centro de la misma, pasando de estar localizada en el sitio 1 a estarlo en el sitio 2.



Para analizar si esta modificación del centro de la curvatura se debía a que se había anulado completamente la interacción de p4 con el sitio 1 y se producía, entonces, por la interacción sólo con el sitio 2 o, si por el contrario, era fruto de la interacción de p4 con ambos sitios, se diseñó un fragmento de DNA que contenía las mismas mutaciones que Mut1, pero que sólo poseía el DNA de  $\phi 29$  correspondiente al sitio 1. En paralelo se ensayó un DNA del mismo tamaño pero correspondiente al DNA del sitio 1 silvestre. Con estos fragmentos de DNA se analizó la capacidad de unión de p4 a concentraciones crecientes de proteína. El resultado, como se puede ver en la figura 18C, fue que se necesitó ocho veces más cantidad de proteína p4 para que se formase el complejo con el DNA mutado y, además, el complejo formado parece más inestable que el del silvestre. Estos datos sugieren que el resultado final de la curvatura que se produce al ensayar el DNA Mut 1 implica la interacción de p4 con ambos sitios de unión, aunque la unión al sitio 1 sea débil e inestable.

Los ensayos con el DNA mutante **Mut 2** mostraron que el fragmento de menor movilidad era el generado con la enzima de restricción *StuI* I, y el obtenido con *Bgl* II era el que presentaba mayor movilidad, al igual que ocurría con la región 1 salvaje (no mostrado). Se dedujo, por tanto, que el centro de curvatura estaba en el mismo lugar o muy próximo a donde se encontraba en el DNA no modificado. El grado de curvatura aparente obtenido con el DNA Mut 2 fue de  $83^\circ$ .



La secuencia del plásmido pBend2-Mut 3 mostró que la orientación del fragmento clonado era opuesta a la del resto de secuencias analizadas; en este caso, se consideró como posición 1 el primer nucleótido obtenido tras la digestión con la enzima *Bgl* II. Cuando se llevaron a cabo los ensayos de permutación circular con este DNA Mut3, se observó que el fragmento de menor movilidad era el obtenido tras la digestión con *Spe* I y el de mayor movilidad el obtenido por el corte de *Bam* HI (figura 19A). Estos datos nos permiten localizar el centro de curvatura en el sitio 1 aunque un poco desplazado de donde se encuentra en el DNA silvestre. El grado de curvatura inducida sobre este DNA Mut3 fue de 89° (figura 19B).

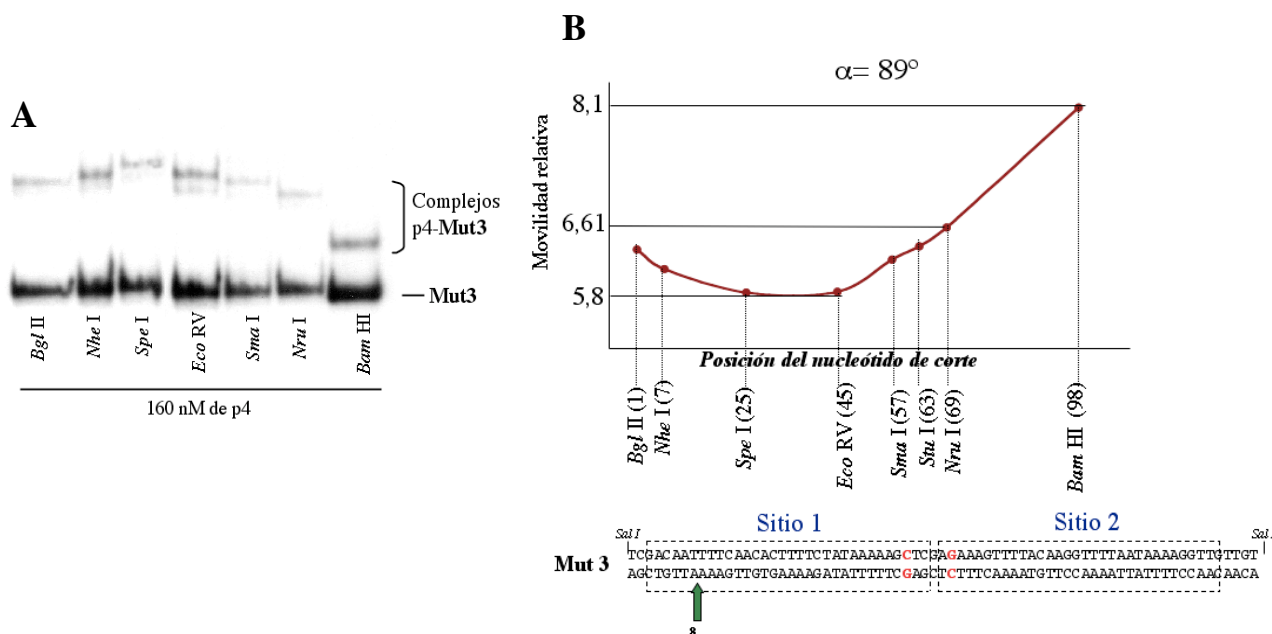


Figura 19: Permutación circular del fragmento Mut3 de la región 1 de  $\phi 29$ . En la autorradiografía del gel de poliacrilamida al 6%, donde se llevó a cabo la electroforesis a 4°C (A), se pueden observar los complejos formados con los 7 fragmentos obtenidos tras la digestión con las enzimas de restricción indicadas y la proteína p4. En la gráfica (B), se muestra el resultado de la extrapolación de los datos del experimento, se observa que la posición del nucleótido respecto al punto de corte es diferente para cada enzima de la que poseían los DNAs ensayados anteriormente, debido a que la orientación del fragmento clonado en el plásmido es la opuesta.

### 3.4) Curvatura inducida por el complejo p4-p6:

Para regular el proceso de transcripción del fago, la proteína p4 forma parte de un complejo nucleoproteico que incluye a la proteína viral temprana p6. Por lo tanto, se decidió estudiar también la conformación del DNA de la región 1 en presencia de ambas proteínas. Para ello, se realizaron ensayos de permutación circular utilizando el plásmido pBend2-wt, que tiene clonada la secuencia de los sitios 1 y 2. Como se puede ver en la figura 20, la movilidad del fragmento digerido con *Bam* HI es menor en presencia de las dos proteínas p4 y p6 (carril c) que cuando se une p4 sola (carril b), o p6 sola (carril a). Toda la proteína p4 del ensayo que está interaccionando con el DNA lo hace en forma de complejo p4-p6-DNA, mientras que la proteína p6, por encontrarse en exceso, da lugar además, a complejos p6-DNA. Tras llevar a cabo las medidas de las distancias relativas, se observó que aunque la migración de los complejos es diferente a la que presentan en presencia de p4 sola, la movilidad relativa de cada fragmento en presencia de p4 y de p6 resultó muy similar a la que se obtuvo en presencia de p4 sola. Este resultado, obviamente, implica que el fragmento que presenta menor movilidad es también el digerido con *Stu* I indicando que el centro de curvatura tiene la misma

posición que cuando se une p4 sola. Cuando se llevaron a cabo los cálculos del grado de curvatura, se obtuvieron  $81^\circ$ , muy próximo a los  $86^\circ$  obtenidos en presencia solo de p4, lo que implica que la conformación del DNA en la región 1 en presencia del complejo regulador p4-p6 es, prácticamente, la misma que la que presenta la proteína p4 sola.

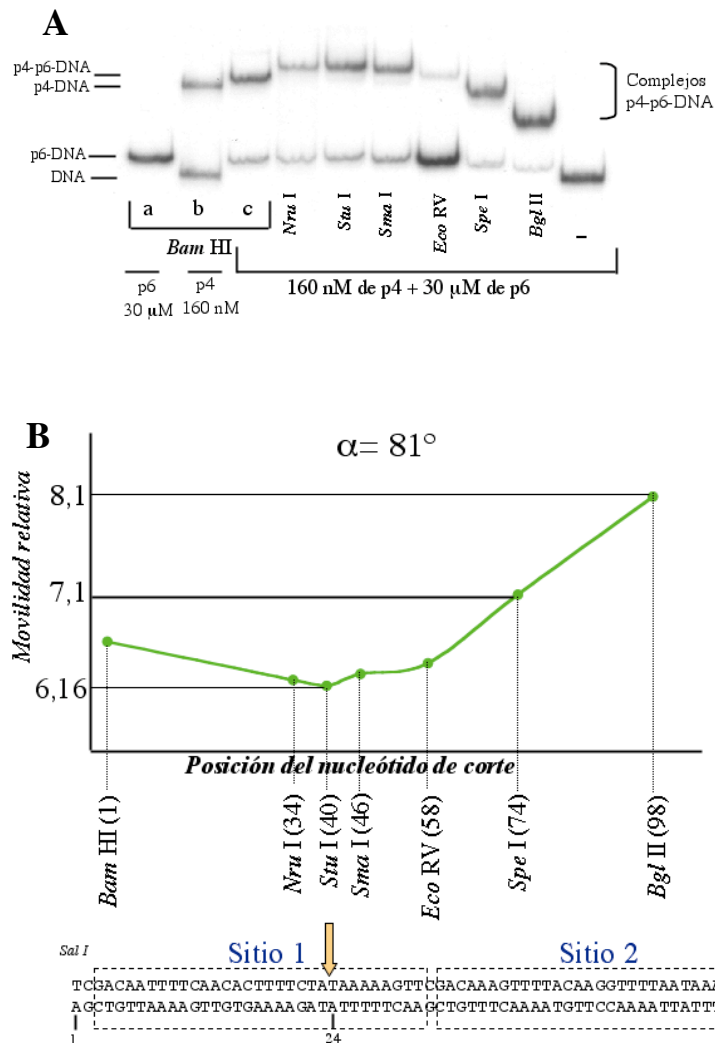


Figura 20: Análisis de la curvatura inducida por la proteína p4 en presencia de la proteína p6. La formación de los complejos de ambas proteínas con los diferentes fragmentos de DNA, procedentes de la digestión del plásmido pBend2-wt, se resolvió por electroforesis a  $4^\circ\text{C}$  en geles al 6% de poliacrilamida (A). En el carril **a** se muestra la movilidad del complejo que forma la p6 con el fragmento digerido con *Bam* HI, en el carril **b** la movilidad del complejo formado por la p4 con el mismo fragmento, y en el carril **c** la migración del complejo formado con ambas proteínas juntas, que es el de menor movilidad. La gráfica (B) muestra el resultado de la extrapolación de las medidas de movilidad relativa de los diferentes fragmentos formando el complejo con las proteínas p4 y p6, y permite comprobar que el centro y grado de curvatura es prácticamente el mismo que se obtiene cuando el complejo con el DNA lo forma la proteína p4 sola.

#### 4.- MUTANTES DE p4 DE UNIÓN AL DNA

En un proyecto de colaboración con el grupo que dirige el Dr. Miquel Coll en el Institut de Biología Molecular de Barcelona (CSIC) y del Institut de Recerca Biomédica, Daniel Badía resolvió la estructura cristalográfica de la proteína p4 y de p4 unida a un fragmento de DNA de 41 pares de bases que contenía el sitio 3 y parte del sitio 4 (Badía *et al.*, 2006).

Los datos obtenidos de la resolución de estas estructuras han sido muy importantes para el desarrollo de nuestro trabajo, tanto para los estudios de la interacción de p4 con el DNA, como para la determinación de las características de sus sitios de unión.

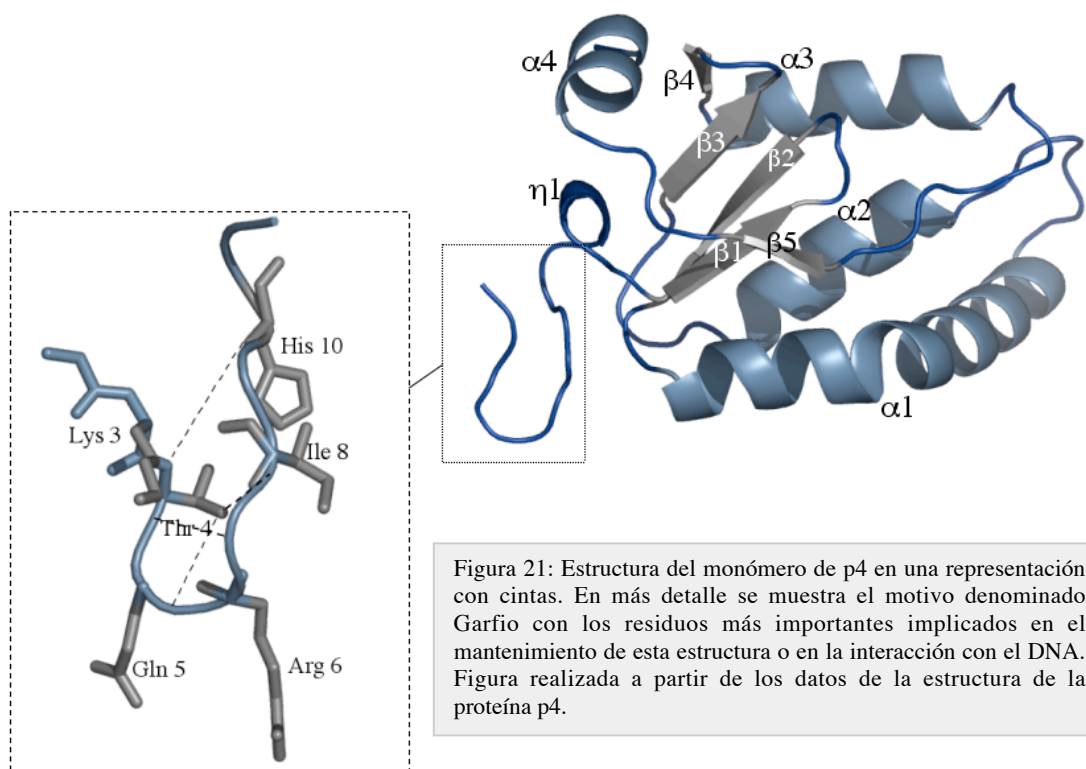
Se resumen a continuación los datos más relevantes extraídos de la estructura y que nos han permitido, entre otros, el diseño de mutantes de p4 en unión al DNA.

##### Monómero de p4:

La estructura secundaria de cada monómero de p4 está formada por cinco cadenas de láminas  $\beta$ -antiparalelas, cuatro  $\alpha$ -hélices y una hélice  $3_{10}$  (figura 21).

En la estructura se observó la presencia de un elemento inusual, en el cual la cadena del polipéptido que va desde la prolina en posición 2 a la glutamina 5 es antiparalela al tramo que va desde la arginina 6 a la asparagina 11 (figura 21). La configuración de este elemento se mantiene por el establecimiento de al menos tres enlaces de hidrógeno entre las cadenas principales y un giro- $\beta$  de tipo I. Esta estructura fue denominada Garfio y es fundamental para el reconocimiento del DNA.

La  $\alpha$ -hélice 4, que está localizada en el extremo carboxilo terminal de la proteína, mostró muy baja densidad en los dos monómeros de la estructura sin DNA, lo cual indica que se trata de una región muy flexible.



### Complejo p4-DNA:

La estructura del complejo mostró que la entidad que se une a un sitio de p4 es un dímero, que contacta con el DNA en tres zonas, y que los dos monómeros se unen al DNA por la misma cara. Cuando se compararon las estructuras de p4 libre o unida al DNA se vio que no existían cambios significativos (figura 22).

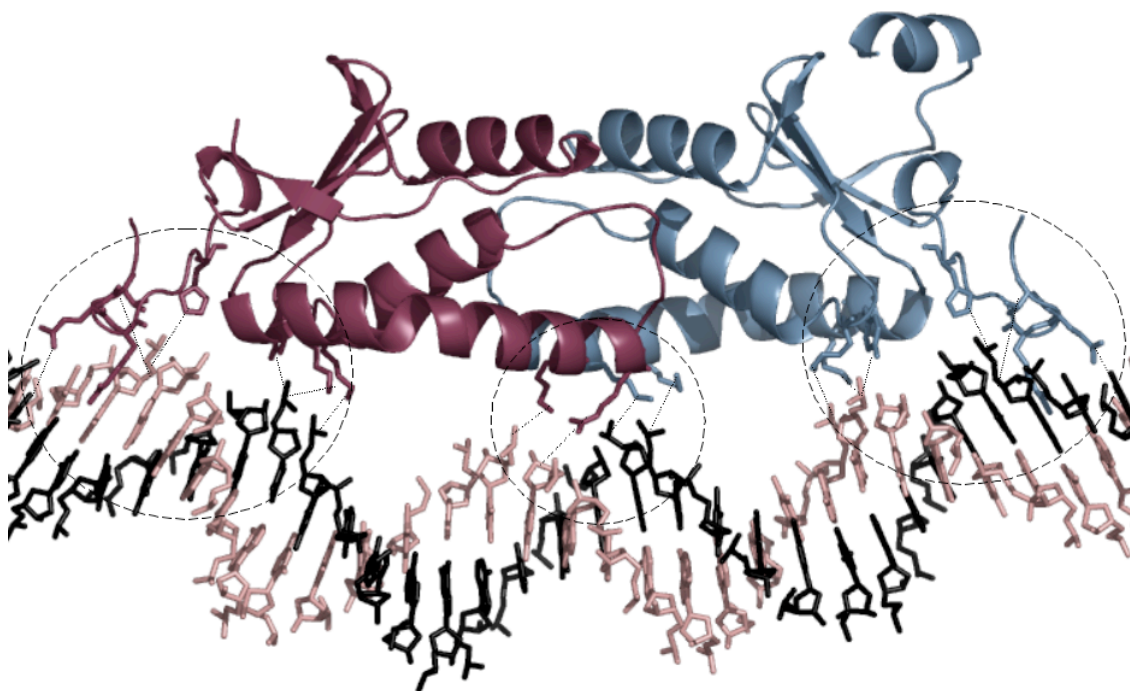


Figura 22: Estructura del dímero de p4 unido al sitio 3. En círculos punteados se indica la localización de las 3 zonas por las que el dímero contacta con el DNA. Se muestran las cadenas laterales de los aminoácidos que interaccionan con el DNA. Figura realizada a partir de los datos de la estructura del complejo p4-DNA.

El DNA de 41 pb que se utilizó para la resolución de la estructura del complejo con la proteína, contiene solo la secuencia completa de un sitio de unión de p4, el sitio 3. El DNA presentó una conformación B con una curvatura de 70° y esta curvatura está dirigida hacia la proteína. En las zonas donde la proteína interacciona con el esqueleto fosfato el surco menor se estrecha; en cambio, el surco mayor es regular a lo largo de la secuencia.

Considerando un sitio de unión de p4 como un fragmento de DNA de 31 pares de bases con una secuencia determinada, se decidió denominar al par de bases central como la posición cero. A partir del cero, hay 15 pares de bases a cada lado que se definieron con números positivos hacia la derecha y negativos hacia la izquierda, en función de la orientación que cada sitio posee en el genoma del virus, estando en la izquierda el promotor A2c y en la derecha el promotor A3 (figura 23B).

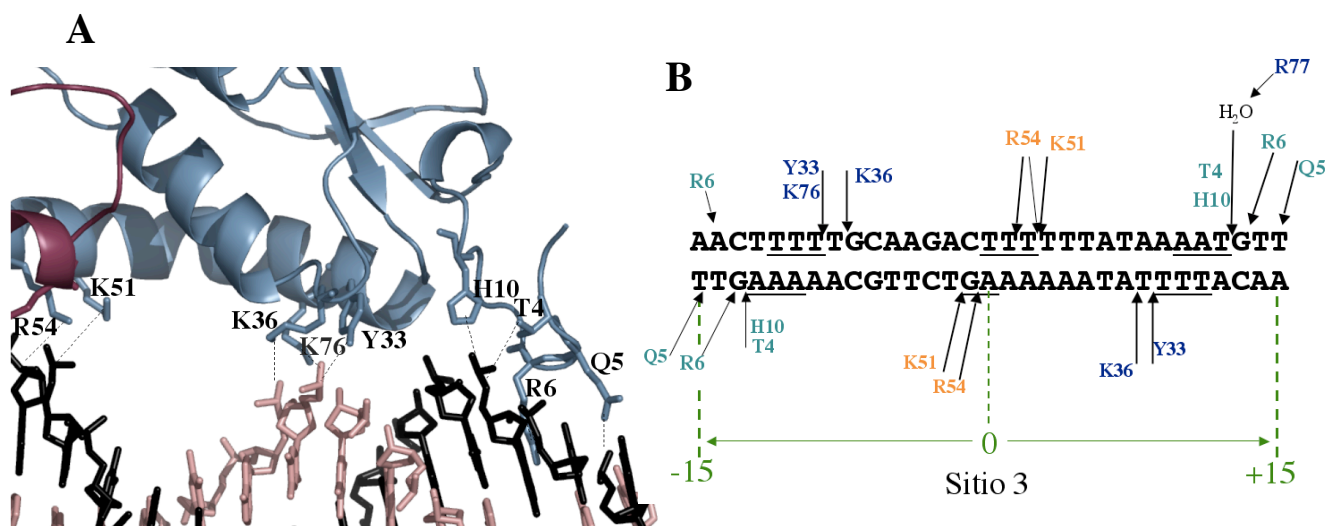


Figura 23: Interacción de p4 con el DNA. En A, se muestran las cadenas laterales de los residuos a través de los que tienen lugar las interacciones de uno de los monómeros con el DNA. En B se muestra un esquema, sobre la secuencia del sitio 3, de las interacciones que establece cada aminoácido del dímero. En la parte inferior de la figura, en verde, se muestra la numeración de los nucleótidos del sitio. Las líneas negras horizontales representan los nucleótidos del sitio 3 protegidos por la unión de p4 a la digestión de radicales hidroxilo que hemos denominado huellas.

La interacción tiene lugar a través de diferentes estructuras secundarias de la proteína:

*El Garfio:* por medio de este elemento tienen lugar los dos únicos contactos que ocurren de forma específica entre un aminoácido y la base nitrogenada de un nucleótido. Se trata del residuo Arg6 de ambos monómeros, que interacciona con las guaninas en posición 13, y de la Gln5 que interacciona con las timinas localizadas en los extremos del sitio 3, en las posiciones 15. Se dan también, dentro de los aminoácidos que forman parte del Garfio, interacciones con el esqueleto del DNA, que son las que establecen la Thr4 y la His10 con los fosfatos de las guaninas en posición 13. Cada garfio se mete dentro de un surco mayor del DNA, siendo la distancia entre los dos Garfios de 75Å.

*La  $\alpha$ -hélice 1:* En su extremo amino se establecen dos interacciones con el esqueleto fosfato del DNA a través de los residuos Lys36 y Tyr33; estos contactos se dan, en ambos monómeros, en las posiciones 7 y 8, respectivamente. En su extremo carboxilo, localizados en la zona de dimerización, dos aminoácidos establecen interacciones con el esqueleto fosfato de la secuencia central del sitio 3. La Lys51 en la posición 2 y la Arg54 que establece en un monómero dos interacciones, en las posiciones +1 y +2, y en el otro solo establece una en la posición -1.

*El bucle de unión de la  $\alpha$ -hélice 2 y la lámina  $\beta$ 3:* en este bucle existen dos residuos implicados en la interacción del dímero con los extremos del sitio. Por un lado la Lys76, que contacta directamente con el fosfato de la timina en posición -8 y por otro la Arg77 de uno de los monómeros que, aunque mediado por una molécula de agua, interacciona con la timina en posición +12.

#### 4.1) Mutantes de p4:

Para estudiar la especificidad de las interacciones y teniendo en cuenta los datos obtenidos tras la resolución de la estructura cristalográfica, se diseñaron mutantes puntuales de la proteína p4 en los aminoácidos implicados en la interacción con el DNA. Se cambiaron los residuos seleccionados por alanina; esto implica que, el cambio introducido disminuye el volumen y la longitud de la cadena lateral y quita cargas, en el caso de que las haya. Podemos dividir en tres grupos los aminoácidos seleccionados para la mutagénesis y posterior análisis.

- En primer lugar, aminoácidos que forman parte del Garfio, que es el elemento más importante en el reconocimiento e interacción con el sitio de unión. Se modificaron cuatro residuos: los dos que establecen interacciones específicas con bases, dando lugar a las proteínas Q5A y R6A y los otros dos aminoácidos que interaccionan con el esqueleto fosfato del DNA, His10 y Thr4 (en verde-azulado en la figura 23B).

- En segundo lugar, se cambiaron a alanina aminoácidos que interaccionan con los extremos de la secuencia del sitio 3, localizados fuera del Garfio, aunque próximos a él y fueron Tyr33, Lys36 y Lys76 (en azul oscuro en la figura 23B).

- Por último, se seleccionaron los dos aminoácidos que establecen interacciones con el DNA en la región central del sitio de unión; se trata de la Lys51 y de la Arg54, localizadas en la región de dimerización de la proteína (en naranja en la figura 23B).

#### 4.2) Unión al DNA:

La elección de los residuos de la proteína que se modificaron, como se ha comentado, se hizo en base a los datos extraídos de la estructura cristalográfica del complejo p4-Sitio 3 (figura 22). Para observar, en un primer examen, la implicación de cada aminoácido en la unión de la proteína al DNA e interpretar así las interacciones observadas en la estructura, se analizó la formación de complejos p4-DNA, utilizando concentraciones crecientes de las proteínas mutadas y un fragmento de DNA con la secuencia del sitio 3. En la figura 24 se muestra el resultado de los ensayos con estas proteínas en la formación del complejo p4-DNA.

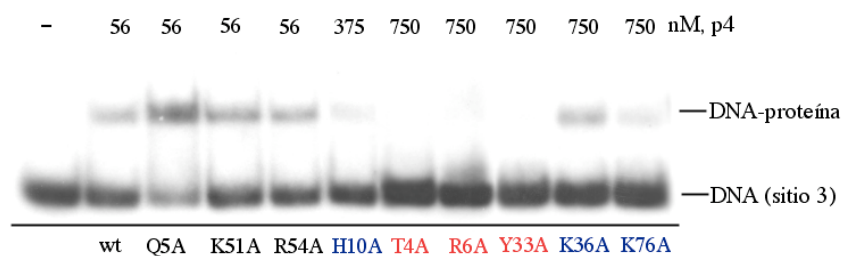


Figura 24: Interacción de las proteínas modificadas de p4 con el DNA del sitio 3. Diferentes concentraciones de las proteínas p4 modificadas se incubaron con un fragmento marcado radiactivamente de 60 pb que contenía la secuencia del sitio 3. La formación de los complejos se resolvió en geles al 6% de poliacrilamida a 4°C. La autorradiografía muestra que tres proteínas (en rojo) no forman ningún complejo con este fragmento de DNA, mientras que en azul se muestran aquellas en las que se requirió mayor cantidad que de la p4 silvestre para formar el complejo. En negro las que poseen un comportamiento igual o mejor que la silvestre.

El comportamiento de las proteínas fue muy diverso. K51A y R54A formaron el complejo con el DNA igual que la proteína p4 silvestre; los residuos modificados en estas proteínas están localizados en la región de dimerización de p4. Por otro lado, hay un grupo de proteínas, H10A, K36A y K76A que, para formar el complejo con el DNA, requirieron entre 8 y 13 veces más cantidad de proteína que la silvestre (figura 24). Sin



embargo, el efecto más drástico en la interacción con el sitio 3 se obtuvo con las proteínas modificadas R6A, Y33A y T4A, que no formaron el complejo con el DNA a ninguna de las concentraciones ensayadas. Por último, hay que destacar el sorprendente comportamiento que mostró la proteína Q5A, ya que, teniendo modificado un aminoácido que establece, en principio, interacciones específicas con las timinas en posición 15 del sitio 3, no solo no está afectada su capacidad para interactuar con el DNA sino que produce, a igual concentración de proteína, aproximadamente el doble de complejos que la proteína p4 salvaje.

La unión al DNA de estas proteínas mutantes también se analizó con otros fragmentos, conteniendo dos sitios de unión, el (3+4) ó el (1+2), o los cuatro sitios de unión de p4; en general, el comportamiento de todos los mutantes fue similar al descrito con el sitio 3, salvo para Y33A y para R6A. Se vio que Y33A daba lugar a la formación de un complejo con aquellos fragmentos que contenían más de un sitio de unión, y R6A fue capaz de interactuar con el DNA cuando el fragmento incluía en su secuencia los sitios 3 y 4, dando lugar, además, a varios complejos con movilidad diferente a la p4 silvestre (no mostrado).

#### 4.3) Regulación de la transcripción:

De los ensayos de unión al DNA pudimos concluir que todas las proteínas mutantes analizadas, salvo T4A, eran capaces de unirse al DNA, con mayor o menor afinidad y/o en función del tipo de DNA del que se tratase. Se analizó entonces la capacidad de estas proteínas para regular la transcripción de los promotores de  $\phi 29$ . La regulación no sólo implica la interacción de la p4 con sus sitios, sino también la correcta formación de complejos nucleoproteicos de la proteína p4 con la proteína p6 viral y con la RNA polimerasa de *B. subtilis*. El objetivo de este análisis era estudiar si estas interacciones de las proteínas p4 mutantes con el DNA eran o no funcionales.

En un fragmento de DNA que contenga la secuencia del genoma de  $\phi 29$  donde están localizados los promotores A2c, A2b y A3, la RNAP de *B. subtilis* es capaz de reconocer e iniciar la transcripción desde los dos promotores tempranos (figura 25A).

Cuando a este sistema se le adiciona la proteína p4 silvestre se produce una ligera represión de los promotores tempranos y la activación del promotor tardío A3. Cuando

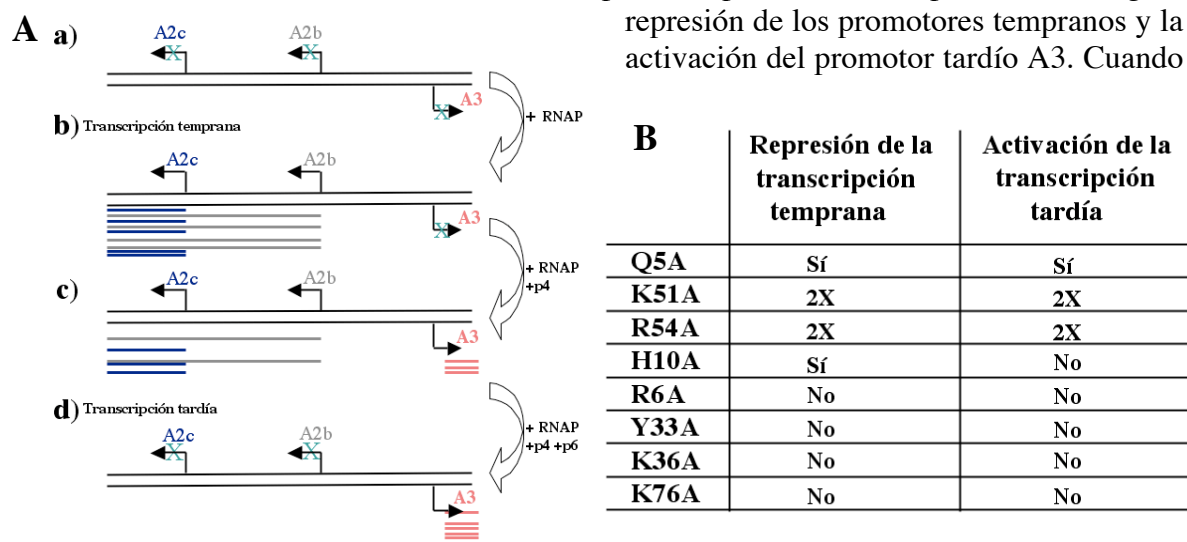


Figura 25: Regulación de la transcripción. (A) Esquema del papel que desempeña cada una de las proteínas implicadas en la regulación de la transcripción truncada *in vitro*. En a) se representa un fragmento de DNA con los promotores A2c, A2b y A3 de  $\phi 29$ . En b) se representa la transcripción temprana que produciría el reconocimiento y la transcripción de los promotores A2c (transcritos azules) y A2b (transcritos grises) por la RNAP. En c) se muestra la represión parcial de los promotores tempranos y la activación del A3 (transcritos rojos) producidas por la presencia de la p4. Finalmente, en d) se muestra un esquema de la transcripción tardía que se produciría en presencia de ambas proteínas reguladoras, p4 y p6, y la simultánea represión de la transcripción temprana. En la tabla B se muestra el resultado obtenido tras llevar a cabo experimentos de transcripción *in vitro* con las proteínas p4 modificadas.

están presentes la proteína p4 y la proteína p6, forman un complejo nucleoproteico que produce una fuerte represión del promotor A2c y del A2b y una mayor activación del A3 (figura 25A). El análisis de la funcionalidad de las proteínas p4 modificadas se llevó a cabo mediante ensayos de transcripción truncada *in vitro* y los resultados obtenidos se recogen en la tabla de la figura 25B. Las proteínas mutantes R6A, Y33A, K36A y K76A son incapaces de regular la transcripción, ya que no reprimen los promotores tempranos ni activan el tardío. Estos ensayos se llevaron a cabo en presencia y ausencia de p6 y con diferentes concentraciones de proteína, no obteniéndose en ningún caso efecto alguno. La proteína mutante H10A posee un comportamiento extraño que implica una fuerte represión de los promotores tempranos, incluso en ausencia de p6, y además no es capaz de activar la transcripción tardía. Para regular la transcripción adecuadamente se requiere el doble de cantidad de las proteínas K51A y R54A con respecto a lo que se necesita cuando se utiliza la p4 salvaje; en cambio, la proteína Q5A regula la transcripción al mismo nivel que la silvestre, utilizando la misma cantidad de proteína.

#### **4.4) Formación de complejos con las proteínas p4 mutantes:**

Con el objetivo de averiguar si la incapacidad de la mayor parte de los mutantes para regular la transcripción se debía simplemente a su deficiente interacción con el DNA o si, además, existía algún defecto adicional en los complejos nucleoproteicos necesarios para la regulación, analizamos la formación de dichos complejos.

##### **4.4.1) Complejos p4-RNAP-DNA:**

Para que la RNAP de *B. subtilis* reconozca el promotor A3 y tenga lugar la activación de la transcripción tardía, es necesario que interaccione con la proteína p4 unida en el sitio 3 (Mencía *et al.*, 1996). Con la intención de estudiar si las proteínas p4 modificadas eran capaces de formar este tipo de complejos con la RNAP, se realizaron ensayos de retraso en la movilidad electroforética del DNA. Para ello, se seleccionó un fragmento de DNA que contenía los sitios 3 y 4 de unión de p4 y el promotor tardío A3. Con este fragmento se ensayó la formación de complejos, utilizando una concentración fija de la RNAP (20 nM) y diferentes dosis de las proteínas mutadas. En las condiciones de este ensayo, la RNAP se une al DNA del promotor A3 en ausencia de p4, puesto que, como se puede ver en la figura 26, en los carriles donde solo hay RNAP se observa una banda de DNA retrasado. Cuando se añade también la proteína p4 salvaje, aumenta el porcentaje de DNA retrasado y la movilidad del fragmento disminuye más, indicando que ambas proteínas están formando un complejo con este DNA. En el panel A de la figura 26 se presenta el resultado de la formación del complejo, con el DNA conteniendo el promotor A3, de la RNAP con los mutantes K51A, R54A y Q5A; las tres proteínas forman el complejo igual que la p4 salvaje. En los paneles B y C se muestra cómo, aunque necesitando mayor cantidad de proteína, los mutantes K36A, H10A, Y33A y K76A son capaces también de dar lugar a un complejo, con la RNAP, de movilidad igual al que produce la p4 silvestre. En el caso de Y33A y K76A no se detecta la formación del complejo p4-DNA en ausencia de la RNAP, lo que implica que estas proteínas modificadas pueden establecer interacciones productivas con la RNAP y como consecuencia, a pesar de su débil interacción con el DNA, son capaces de formar el complejo. En el panel D de la misma figura se presenta el resultado obtenido con la proteína R6A; se puede observar que las cantidades de p4 mutada que se requirieron fueron manifiestamente mayores que de las demás proteínas ensayadas. En ausencia de la RNAP, la proteína R6A da lugar a varios complejos con el DNA, apareciendo bandas en diferentes posiciones y ninguna coincidente con la que produce la p4 silvestre; en cambio, en presencia de la RNAP, la movilidad del complejo formado es la misma que



con la p4 sin modificar. Cuando se realizó el ensayo con la proteína T4A, como se puede ver en el panel E, el mutante no forma ningún complejo con el DNA ni en presencia ni en ausencia de la RNAP. Este resultado, unido a los anteriores obtenidos con esta proteína, sugieren que la cadena lateral de la treonina tiene un papel fundamental en la correcta unión de p4 a sus sitios, ya que la modificación introducida implica la pérdida total de la interacción de p4 con el DNA y, además, la presencia de la RNAP no es suficiente para recuperarla.

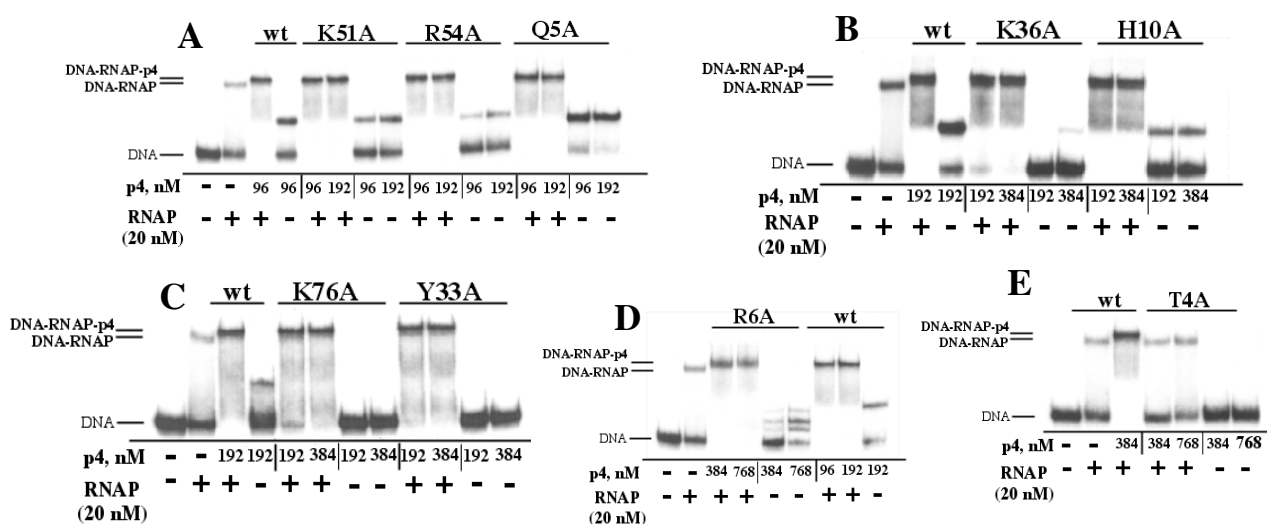


Figura 26: Formación de los complejos p4-RNAP-DNA. Los complejos formados tras la incubación con las diferentes proteínas p4 modificadas y el fragmento de DNA marcado radiactivamente, que contiene el promotor tardío A3 y los sitios 3 y 4, se resolvieron mediante electroforesis a 4°C en geles nativos al 4% de poliacrilamida. En todos los paneles se presenta el comportamiento de la proteína salvaje a las concentraciones indicadas en cada caso. El panel A muestra el resultado positivo de la formación de los complejos de las proteínas K51A, R54A y Q5A. En los paneles B, C y D, se muestra el requerimiento de mayor concentración de proteína para los mutantes K36A, H10A, K76A, Y33A y R6A. Y por último, en el panel E, se recoge el resultado negativo en la formación del complejo que se obtuvo con varias dosis de la proteína T4A.

Puesto que el resto de los mutantes eran capaces de formar el complejo con la RNAP y sabiendo que cinco de ellos no eran funcionales, decidimos estudiar cómo eran estos complejos mediante ensayos de protección del DNA a la digestión con DNasa I, que permiten analizar el patrón que genera el complejo. Para llevar a cabo estos experimentos se utilizó un fragmento de DNA que contenía los cuatro sitios de unión de la proteína p4 (desde la posición +77 del A2c a la +29 del A3).

En la figura 27 se presenta el resultado obtenido en estos ensayos con las proteínas p4 mutantes, salvo con H10A que se analizó por separado (figura 29).

La unión de la RNAP al promotor A2c se caracteriza por la aparición de una hipersensibilidad en la posición -37 y una protección que abarca la secuencia comprendida entre las posiciones -55 y +17 respecto al A2c (comparar carriles a y b; figura 27). La unión de p4 a sus regiones 1 y 2 en el DNA produce hipersensibilidades y protecciones que abarcan desde la posición -20 a la -60 y desde la -122 a la -170 del A2c, respectivamente (comparar carriles a y c, figura 27).

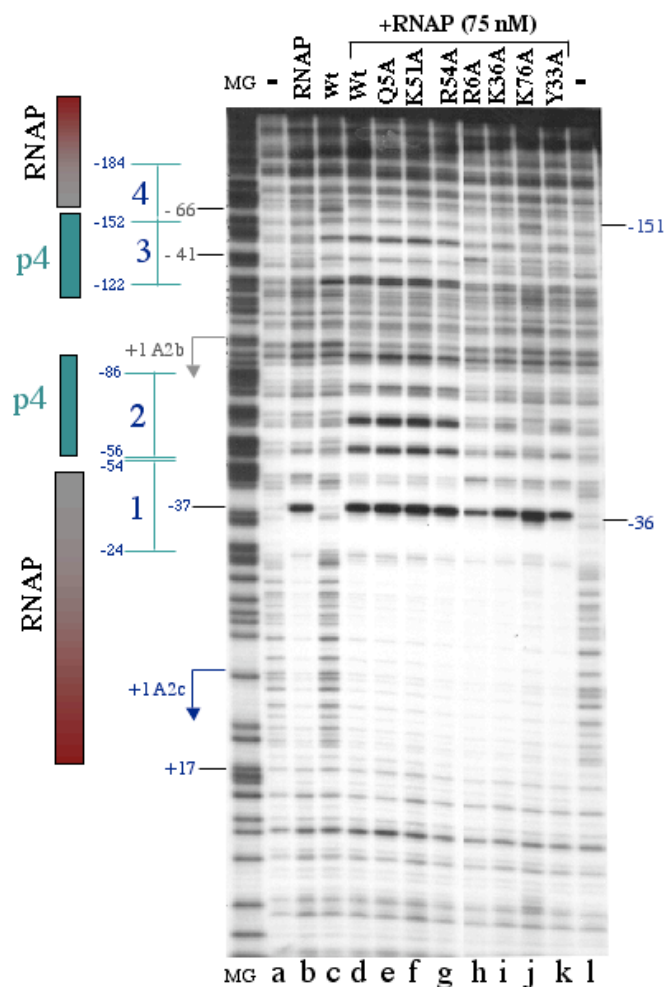


Figura 27: Complejos p4-RNAP-DNA. Un fragmento de 325 pb del genoma de  $\phi 29$  se marcó radiactivamente en el extremo 5' de la posición +77 del promotor A2c. Se incubó con las diferentes proteínas para formar los correspondientes complejos. Las concentraciones de proteínas utilizadas fueron: 75 nM de RNAP, 812 nM de p4 salvaje y de los mutantes Q5A, K51A y R54A; 1,3  $\mu$ M de R6A y de K36A y 1,6  $\mu$ M de Y33A y K76A. El DNA se digirió con la enzima DNasa I y los fragmentos obtenidos se resolvieron en geles desnaturalizantes al 6% de poliacrilamida-8M urea. La autorradiografía muestra (carril b) el patrón de bandas que genera la unión de la RNAP en el DNA, (carril c) la unión de la p4 salvaje y el patrón producido por ambas proteínas (carril d); la formación del complejo con los mutantes se puede ver entre los carriles e-k. Los números en azul indican la posición de los nucleótidos respecto al promotor A2c y los números en gris respecto al A2b. Se muestra la posición de ambos promotores y de los cuatro sitios de unión de p4. El carril MG, es el resultado de la secuenciación del fragmento por el método de Maxam y Gilbert (materiales y métodos). En el margen izquierdo se muestra un dibujo que esquematiza la localización de las proteínas en el complejo con la p4 salvaje.

En presencia de ambas proteínas, RNAP y p4 (carril d), la unión de la p4 al sitio 2 produce la aparición de tres fuertes hipersensibilidades y protecciones entre ellas.

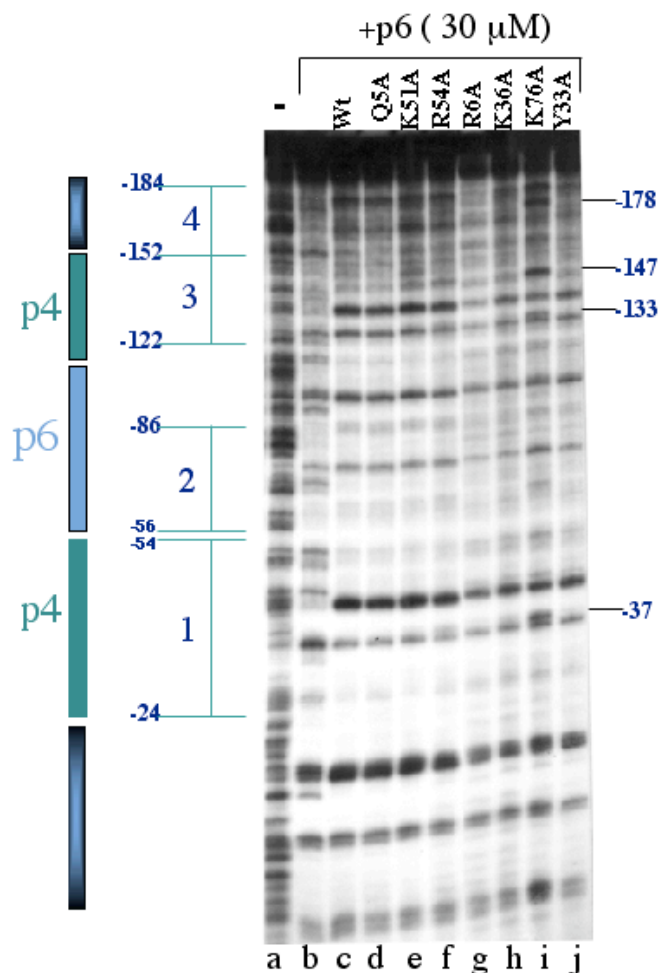
Los mutantes Q5A, K51A y R54A dan lugar, en presencia de la RNAP, a un patrón de bandas idéntico al que produce la p4 silvestre (líneas d, e, f y g). En el caso de los mutantes R6A, K36A, K76A e Y33A los complejos formados tienen un patrón similar al que presenta la RNAP sola (comparar línea b con h, i, j y k) pero existen algunas diferencias, principalmente en las secuencias correspondientes a los sitios de unión de p4 (figura 27).

Estos resultados indican, al igual que en el apartado anterior, que las proteínas K36A, Y33A, K76A y R6A, forman complejos con la RNAP en el DNA de los promotores de  $\phi 29$ . Sin embargo, estos complejos parecen tener una conformación diferente a la que presenta la p4 silvestre, pudiendo ser la formación inadecuada del complejo lo que haría a estos mutantes incapaces de regular la transcripción.

#### 4.4.2) Complejos p4-p6-DNA:

Para que pueda tener lugar la correcta regulación de los promotores de la región central del genoma, A2c, A2b y A3, es necesaria la formación en esta secuencia de un complejo nucleoproteico con las proteínas p4 y p6 del fago. Estas dos proteínas cooperan para formar con el DNA un complejo que permita la activación de la transcripción tardía mediante la estabilización de la RNAP en el promotor A3 y

simultáneamente la represión de los dos promotores tempranos impidiendo la estabilización de la RNAP en los mismos. Para estudiar si las proteínas mutantes de p4 eran capaces de formar estos complejos con la proteína p6, se llevaron a cabo dos tipos de experimentos. En primer lugar, ensayos de retraso de la movilidad electroforética del DNA, donde los mutantes Q5A, K51A, R54A, K36A y H10A presentaron un complejo igual al de la proteína salvaje y los mutantes K76A, Y33A y R6A y T4A, o no formaban el complejo o éste era inestable (no mostrado). En segundo lugar, para estudiar cómo



eran estos complejos, se realizaron ensayos de protección del DNA a la digestión con DNasa I. En el complejo p4-p6 en un DNA conteniendo los cuatro sitios de unión, la proteína p4 silvestre permanece unida a los sitios 1 y 3, mientras que la proteína p6 se une a lo largo de la secuencia comprendida entre ambos sitios (Camacho y Salas, 2004). La formación de este complejo con la proteína p4 silvestre se puede ver en el carril c de la figura 28; la dosis de p6 utilizada en este ensayo provoca su unión también por delante del sitio 1.

Figura 28: Análisis de los complejos p4-p6-DNA. El fragmento de DNA marcado en la posición +77 del A2c se incubó con una concentración 30  $\mu$ M de p6 y con las diferentes proteínas p4 (812 nM de Q5A, K51A, R54A y p4 salvaje, 1,3  $\mu$ M de K36 y R6A y con 1,6  $\mu$ M de Y33A y K76A). Tras la digestión con la DNasa I, se resolvieron los fragmentos obtenidos en un gel desnaturalizante al 6% de poliacrilamida-8M urea. Las posiciones señaladas son con respecto al promotor A2c. En el margen izquierdo de la figura se muestra un esquema de la posición de las proteínas p4 salvaje y p6 en el complejo.

Los mutantes presentan un patrón similar al de la proteína salvaje; en el caso de Q5A, K51A, R54A el patrón es idéntico, mientras que en el caso de R6A, Y33A y K36A el patrón es igual pero tanto las hipersensibilidades como las protecciones son más tenues, lo que indica que el número de moléculas de DNA en las que se está formando el complejo p4-p6 es menor. En el caso de K76A, el patrón es como el de estos últimos mutantes salvo por la aparición de cuatro hipersensibilidades adicionales. La apariencia de los complejos p4-p6 que forman los mutantes R6A, Y33A y K36A en el DNA es la de complejos débiles que probablemente no favorezcan la estabilidad del sistema y, por tanto, tampoco la correcta regulación de la transcripción.

En el caso de H10A en presencia de la proteína p6 (figura 29, carriles c y d), se observa la unión de forma débil de la proteína mutante sólo al sitio 3; además se pueden apreciar hipersensibilidades debidas a la unión de p6, no formándose el complejo como con la p4 silvestre (figura 28). En cambio, con la RNAP (carriles e y f) la proteína H10A forma un complejo igual al de la p4 salvaje, sugiriendo que el defecto en regulación de la transcripción observado podría deberse a una interacción inadecuada con la RNAP.

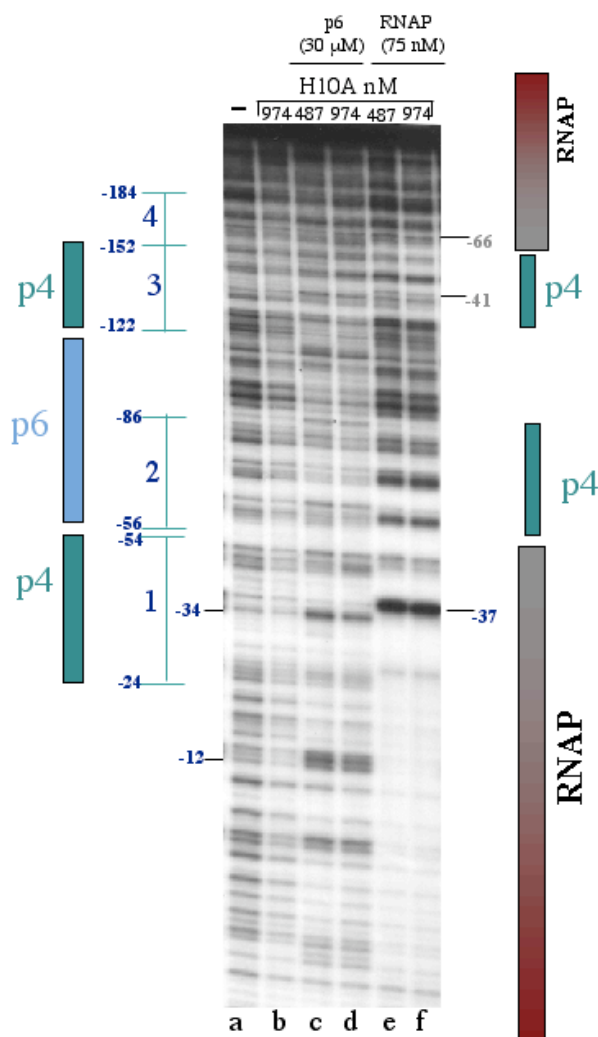


Figura 29: Complejos H10A-RNAP-DNA y H10A-p6-DNA. El DNA marcado radiactivamente en la posición +77 del A2c se incubó con las diferentes concentraciones de las proteínas que se indican en la figura. Se resolvieron los fragmentos obtenidos tras la digestión con DNasa I en un gel desnaturalizante al 6% poliacrilamida-8M urea. En la figura se muestran en azul las posiciones desde el promotor A2c y en gris desde el A2b. En los márgenes de la figura se representa la posición de las proteínas en los complejos que forma la proteína p4 salvaje con la RNAP y con p6.

#### 4.5) Capacidad de inducir una curvatura en el DNA:

Uno de los aspectos que caracterizan la unión de la proteína p4 al DNA es su capacidad para curvarlo. Nos propusimos estudiar si los mutantes que estábamos analizando y que poseían mutaciones en aminoácidos implicados en la interacción de la proteína con el DNA eran capaces de curvar el DNA, y en caso de producirse, estudiar cómo era esta curvatura en cuanto a la localización de su centro y a los grados que generaba. Para ello, se llevaron a cabo ensayos de permutación circular, utilizando un fragmento de DNA que contenía la secuencia de los sitios 1 y 2 (plásmido pBend2-wt). En la figura 30 se puede ver el resultado de los geles y las gráficas correspondientes a los ensayos con los mutantes Q5A, K51A y R54A.

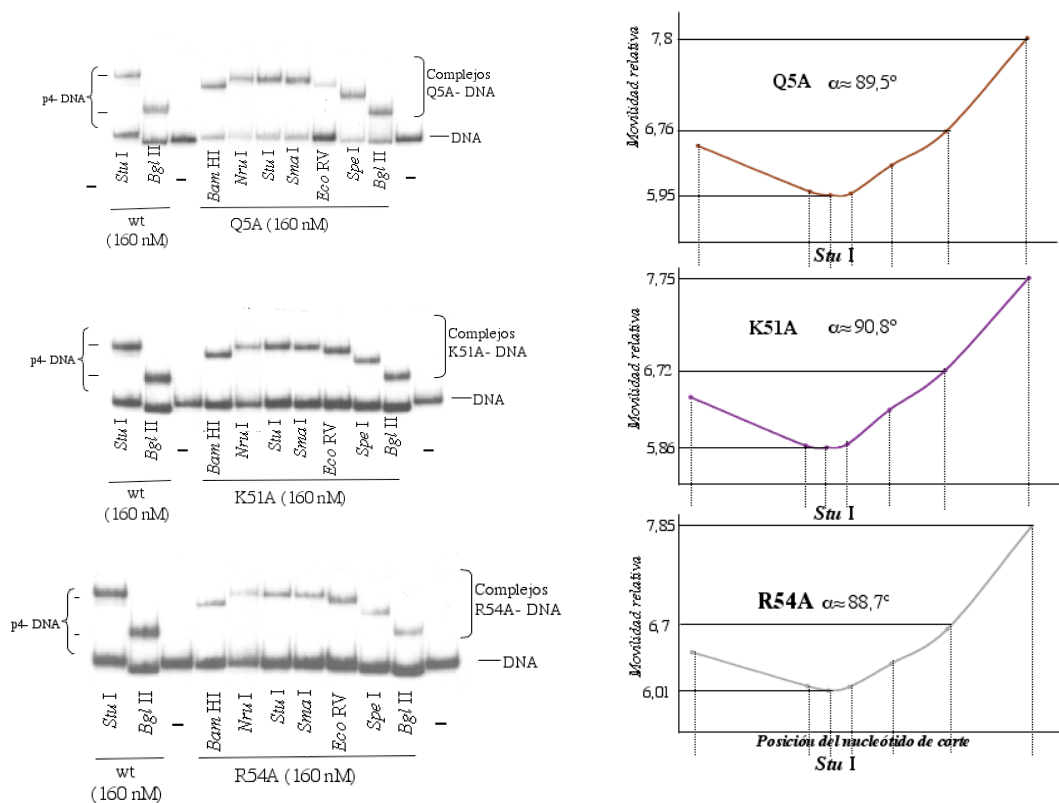
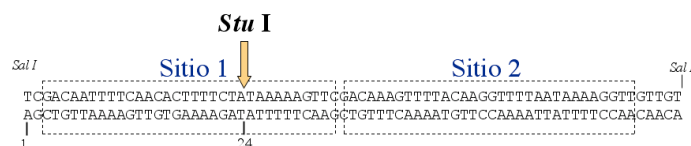


Figura 30: Estudio de la curvatura inducida en la región 1 por las proteínas p4 modificadas Q5A, K51A y R54A. Se muestran las autorradiografías de los ensayos de permutación circular realizados; se presentan también las gráficas resultado de la extrapolación de las medidas de movilidad relativa, así como la localización del centro de curvatura y los grados de la misma.



Debido a que estos mutantes se unen de forma estable a la secuencia de DNA que contiene la región 1 se pudo analizar con detalle la curvatura inducida tras su unión al DNA. El resultado mostró que todos ellos inducían una curvatura igual a la generada por la proteína salvaje, es decir, cercana a los  $86^\circ$  y centrada en el sitio 1 (aproximadamente en el nucleótido 24).

Los mutantes H10A y K36A se unen muy débilmente a la región 1 del DNA, para poder realizar una estimación aproximada de las modificaciones conformacionales generadas en el DNA por estas proteínas mutantes, se procedió por comparación entre la movilidad relativa producida por el complejo formado con H10A y K36A y alguno de los fragmentos con la región 1 permutada, con respecto a la movilidad relativa presentada por estos mismos fragmentos formando el complejo con la proteína p4 silvestre. En la figura 31 se muestra el resultado del ensayo con los fragmentos digeridos con las enzimas *Bgl* II y *Stu* I, que son los que poseen, con la proteína silvestre, mayor y menor movilidad, respectivamente. La posición en la que se localizan las bandas del DNA formando complejo con K36A y con H10A es la misma que la de las demás proteínas analizadas. Tras hacer las medidas de movilidad relativa, los valores

fueron prácticamente iguales para todas. De este y de otros ensayos similares pudimos deducir que ambas proteínas inducen una curvatura cuando se unen a la región 1, que ésta es de aproximadamente 86° y que está centrada en el mismo lugar o muy próximo a donde la inducen la proteína silvestre y los mutantes Q5A, K51A y R54A.

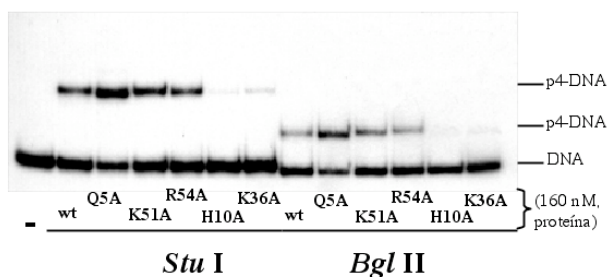


Figura 31: Análisis de la curvatura inducida por las proteínas H10A y K36A, mediante comparación de la movilidad relativa que se genera al interaccionar la p4 salvaje y otras proteínas mutantes con los fragmentos de la región 1 digeridos con *Bgl II* y *Stu I*

La inducción de la curvatura por las proteínas K76A, Y33A y R6A, no se pudo analizar, porque no se unen de forma detectable por esta técnica a un DNA que contenga sólo los sitios 1 y 2. Por lo tanto, todos los mutantes ensayados son capaces de curvar el DNA, pero lo más interesante es que todos ellos dan lugar a una curvatura centrada en el sitio 1, y que el grado de curvatura alcanzado en todos los casos es muy similar. La conclusión a la que nos permitieron llegar estos ensayos fue que la proteína p4 produce una curvatura cuando se une al DNA conteniendo los sitios 1+2 que es siempre alrededor de los 86° siendo ésta, probablemente, una característica funcional de la p4.



## 5.- ESTUDIO DE LOS DETERMINANTES DE ESPECIFICIDAD DE LA SECUENCIA DE UN SITIO DE UNIÓN DE p4

En el apartado 1 de este trabajo se describió la caracterización del sitio mínimo de unión de p4. Este sitio contiene la secuencia correspondiente a tres huellas consecutivas, obtenidas mediante ensayos de protección a la ruptura química del DNA

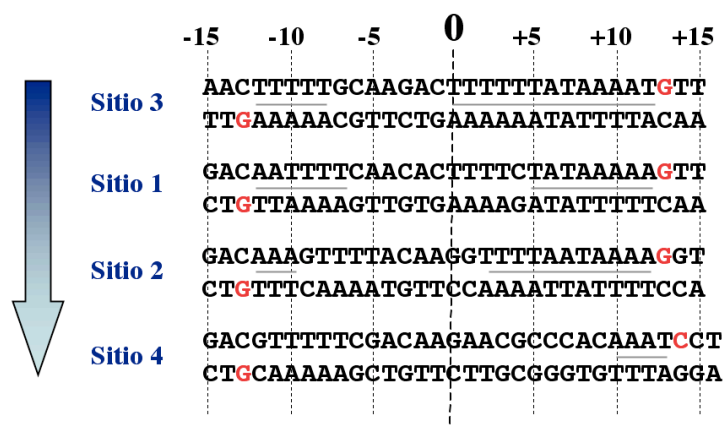


Figura 32: Secuencia de los cuatro sitios de unión de p4. En la figura se señala la numeración de los nucleótidos de un sitio de unión, de cinco en cinco, a partir de la posición central (cero). Se marcan en rojo las guaninas en posición 13 con las que interaccionan las Arg 6 del dímero de p4. El sitio 4 posee un nucleótido más en la secuencia y carece de la guanina correspondiente a la posición +13. Con líneas grises se marcan las secuencias de adeninas y timinas consecutivas de cada sitio. El sentido de la flecha azul indica el orden decreciente de la afinidad de p4 por sus sitios.

por radicales hidroxilo (figura 13), y además es coincidente con lo que llamamos una repetición invertida en dicha secuencia. También se describió que existen cuatro sitios de unión de la proteína y que p4 presenta diferente afinidad de unión por cada uno de ellos. Si analizamos la secuencia de los sitios que se recoge en la figura 32 podemos observar que se trata de un fragmento de DNA de 31 pb especialmente rico en adeninas y timinas.

Mediante ensayos de interferencia por metilación se había descrito la importancia en la interacción

de p4 con el DNA de las dos guaninas, separadas por 25 pares de bases, presentes tanto en el sitio 3 (Barthelemy y Salas, 1989), como en los sitios 1 y 2 (Monsalve *et al.*, 1998); estas guaninas son las que están localizadas en las posiciones +/- 13 (en rojo, figura 33). La resolución de la estructura de la proteína p4 unida al DNA permitió observar cómo se producía esta interacción en el sitio 3 (figura 22) y aportó mucha información que permitió especular sobre cuáles podrían ser los determinantes más importantes que caracterizan a un sitio de unión de la proteína p4.

Por una parte, se observó que sólo dos aminoácidos de cada monómero de p4 establecían interacciones específicas con bases nitrogenadas del DNA (figura 23). Una de las interacciones específicas se produce entre las argininas 6 del dímero con las guaninas +13 y -13 del sitio 3, lo cual confirma la importancia que se había dado a estas guaninas. Estas bases están presentes en todos los sitios de unión de p4 salvo en el sitio 4 que es el de menor afinidad y que solo posee una de ellas (figura 32). La otra interacción específica se produce entre las glutaminas 5 de la proteína y las timinas en posición 15 del DNA; teniendo en cuenta que el sitio 3 es por el que p4 presenta mayor afinidad y que es el único que posee las dos timinas en posición 15, pensamos que la presencia de ambas bases en estas posiciones podría ser una de las características determinantes de la mayor afinidad que p4 presenta por el sitio 3 con respecto a los demás sitios. El resto de interacciones se producen entre los aminoácidos y el esqueleto fosfato del DNA; como se puede ver en la figura 23, las tres zonas del DNA donde tienen lugar este tipo de interacciones son prácticamente coincidentes con las posiciones

que resultan protegidas en los ensayos con radicales hidroxilo, y que previamente se han denominado huellas.

Por otra parte, se observó que el fragmento de DNA con el que está cristalizada la proteína y que posee el sitio 3 y parte del sitio 4, alcanza un grado de curvatura de  $70^\circ$ , por lo que debe ser flexible. Tomando como referencia el sitio 3, se puede observar que a partir de las guaninas en posición 13 y hacia el centro del sitio, existen una serie de adeninas y timinas consecutivas en cada lado; en el resto de los sitios también aparecen estos grupos de A/T aunque o bien su longitud es menor o están interrumpidos (líneas grises, figura 32). Teniendo en cuenta estas observaciones, supusimos que la alta presencia de bases A/T en la secuencia podría estar favoreciendo la formación de la curvatura que induce p4 en el DNA y ser, por tanto, una característica importante que confiere al sitio mayor afinidad en su interacción con la proteína.

Partiendo de estos antecedentes y con la intención de averiguar si se trataba de factores importantes que influían en la afinidad de p4 por sus sitios, se realizaron modificaciones en la secuencia de los sitios 1, 2 y 4, con el fin de aumentar dicha afinidad de unión.

### 5.1) Modificaciones en el sitio 1:

La afinidad relativa que presenta la proteína p4 por el sitio 1 es dos veces menor que la que presenta por el sitio 3 (figura 14). Un análisis comparativo de las secuencias de ambos sitios mostró que el sitio 1 carece de la timina en posición -15, y que el grupo de A/T de la mitad derecha del sitio está interrumpido por una citosina en posición +4. Diseñamos dos nuevos “sitios 1” (figura 33), el **Sit 1A** que posee la secuencia del sitio 1 pero tiene modificada la citosina de la posición -15 por una timina, y el **Sit 1B** cuya citosina en posición +4 fue cambiada por una timina y como resultado, el grupo de A/T de la mitad derecha del sitio 1 alcanza la longitud que tiene en el sitio 3, extendiéndose hasta la posición +1. Se ensayó la unión de la proteína p4 a estos fragmentos modificados y se comparó con los complejos formados con el sitio 1 salvaje. Como se puede ver en la figura 33, ninguna de las modificaciones realizadas aumentó significativamente la afinidad de la proteína p4 por el sitio 1. En contra de lo esperado, ni la presencia de una timina en posición -15 en un sitio que carece de ella, ni la prolongación de la serie de residuos de A/T de 8 a 13 pares de bases consecutivos, aumenta la afinidad de unión de la proteína p4.

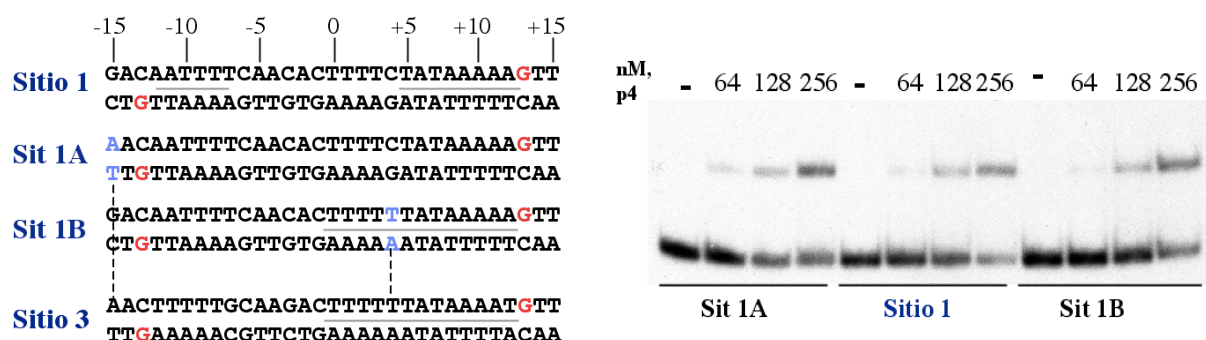


Figura 33: Ensayo de interacción de la proteína p4 con el sitio 1 salvaje y los sitios 1 modificados. Se muestran en azul los nucleótidos cambiados y se compara con la secuencia del sitio 3. Se presenta la autorradiografía obtenida tras analizar los complejos p4-DNA en electroforesis a  $4^\circ\text{C}$ , en geles nativos al 6% de poliacrilamida.



### 5.2) Modificaciones en el sitio 2:

La afinidad de la proteína p4 por el sitio 2 es muy baja, siendo el porcentaje de DNA formando el complejo cinco veces menor que en el caso del sitio 1 (figura 14). Cuando se comparó la secuencia del sitio 2 con la del sitio 3, que usamos como modelo, se observó que aunque la longitud del grupo de A/T de la mitad derecha del sitio era similar, diferían mucho en el resto de la secuencia. El sitio 2 carece de la timina en posición -15, el grupo de A/T de la parte izquierda está interrumpido en la posición -9 y por último, en las posiciones +1 y 0 existen pares de bases G/C que en el sitio 3 son pares de bases A/T. Puesto que el efecto de la timina en posición -15 ya se analizó sobre el sitio 1, nos propusimos estudiar mediante modificaciones en el sitio 2, lo que ocurría con las otras diferencias de las secuencias de los sitios. Se diseñó el mutante **Sit 2B**, en el que se modificó el par de bases G/C por A/T en posición -9 prolongándose el grupo de A/T de la mitad izquierda del sitio 2, alcanzando mayor longitud, incluso, que el sitio 3. Y, por otro lado, diseñamos un mutante que se denominó **Sit 2A** en el que se modificaron las posiciones 0 y +1 por pares de bases A/T. Se analizó la capacidad de la proteína p4 para interaccionar con estos “sitios 2” modificados, mostrándose el resultado en la figura 34. El cambio que implicaba un aumento en la longitud del grupo A/T de la mitad izquierda del sitio, **Sit 2B**, no mejoró la unión de la proteína al DNA. Cuando se modificó la zona central, **Sit 2A**, se observó un aumento de aproximadamente 8 veces en la afinidad de la proteína p4 por el sitio 2, alcanzando, prácticamente, la afinidad que tiene p4 por el sitio 3. La explicación de este aumento en la afinidad de la proteína podría encontrarse en el hecho de que los residuos Lys51 y Arg54 de la región de dimerización de la proteína p4 interaccionan precisamente con el esqueleto fosfato de las bases localizadas en las posiciones 0 y +1 (figura 23).

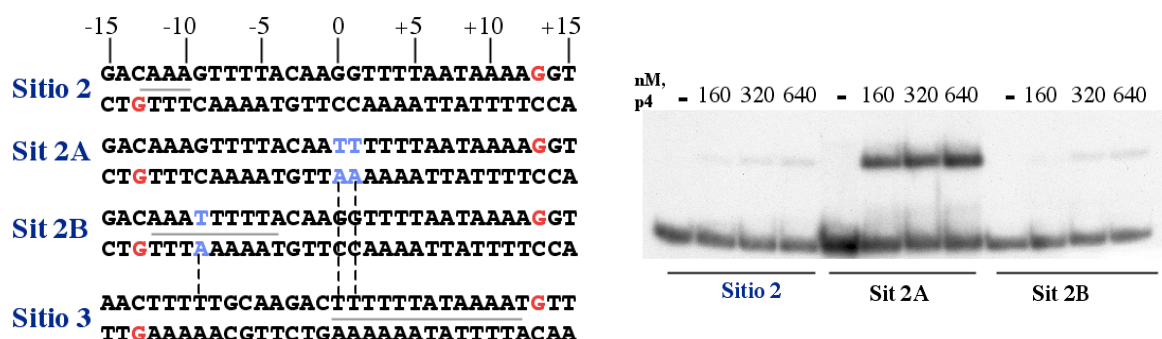


Figura 34: Ensayo de interacción de la proteína p4 con los mutantes del sitio 2. Se muestran en azul los nucleótidos cambiados y se comparan con la secuencia del sitio 3. Se presenta la autorradiografía obtenida tras analizar los complejos p4-DNA en electroforesis a 4°C, en geles nativos al 6% de poliacrilamida.

### 5.3) Modificaciones en el sitio 4:

El sitio 4 es por el que la proteína presenta menor afinidad; de hecho, no se ha podido detectar la unión de p4 a este sitio de forma individual. Las evidencias de que esta unión existe, siempre se obtuvieron en presencia del sitio 3. El sitio 4 carece de una de las guaninas en posición 13 que se han definido como críticas para la unión de p4. Otra característica de los sitios de unión de p4 es su alto porcentaje de pares de bases de adeninas y timinas (siendo 80% en el sitio 3, 77,5% en el sitio 1 y 74% en el sitio 2); el sitio 4 posee sólo un 53%, o lo que es lo mismo, tiene mayor contenido en pares de bases G/C que podría estar disminuyendo la flexibilidad del DNA y perjudicando su

capacidad para curvarse. Además, hay que destacar que el sitio 4 posee un par de bases más que el resto de los sitios. Se llevaron a cabo dos tipos de modificaciones para intentar aumentar la afinidad de unión de p4 por este sitio. En primer lugar, se modificó la citosina en la correspondiente posición +13 por una guanina (**Sit 4A**) siendo, en este caso, la distancia entre las guaninas de 26 pares de bases en lugar de 25 como en el resto de los sitios (**figura 35**). Por otro lado, con la intención de aumentar la longitud del grupo de A/T del extremo derecho del sitio 4 y, al mismo tiempo, el porcentaje de este tipo de bases en el sitio 4 (de 53% a 72%), se cambiaron las bases comprendidas entre las posiciones +8 y +2 por las siete que posee el sitio 3 en esta localización, **Sit 4B**. El análisis de la interacción de p4 con el DNA de estos “sitios 4” modificados se realizó del mismo modo que en los casos anteriores (**figura 35**). El resultado mostró que la incorporación de una guanina en posición 13, **Sit 4A**, produce un aumento de la afinidad de p4 por el sitio y se hace detectable, por primera vez, la unión de forma individual al sitio 4. Esto indica que la interacción es lo suficientemente estable como para resistir las condiciones a las que se somete en la electroforesis. Así se confirma la importancia de la presencia de las guaninas en los extremos de los sitios de unión de p4, aunque el sitio posea un par de bases más y un alto porcentaje de pares de bases G/C; de todos modos, la cantidad de **Sit 4A** que forma el complejo con la proteína p4 es muy pequeña y además no hay linealidad en la formación del complejo p4-DNA, sugiriendo que la afinidad de la proteína por este DNA es muy baja. La modificación de los 7 pares de bases que dan lugar al **Sit 4B** no produce un aumento de la afinidad suficiente como para ser detectable la interacción con p4 por esta técnica.

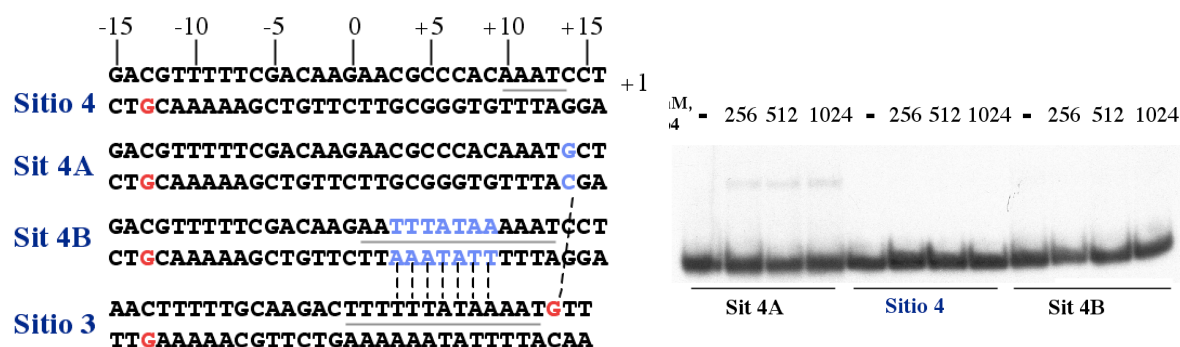


Figura 35: Ensayo de interacción de la proteína p4 con los DNAs correspondientes al sitio 4, Sit 4A y Sit 4B. Se muestran en azul los nucleótidos cambiados con respecto al sitio silvestre y se comparan con la secuencia del sitio 3. Se presenta la autorradiografía obtenida de la electroforesis a 4°C, en geles nativos al 6% de poliacrilamida.

La conclusión principal que podemos sacar de estos resultados es que existen dos factores que afectan muy positivamente a la afinidad de unión que la proteína p4 tiene por sus sitios. Uno de ellos es la presencia de las guaninas que interaccionan específicamente con las Arg 6 del dímero. El otro factor importante es la existencia de pares de bases A/T en las posiciones centrales del sitio (+1 y 0), cuyos fosfatos establecen interacciones con otros residuos de la proteína, Lys51 y Arg56; estas bases A/T están conservadas en los dos sitios por los que p4 tiene mayor afinidad, el 1 y el 3. Otras conclusiones a las que nos permiten llegar estos resultados son que ni la longitud de los grupos de adeninas y timinas presentes en ambas mitades del sitio, ni las timinas en los extremos con las que interacciona la Gln5 de p4, son factores que influyan en la afinidad que p4 tiene por sus sitios de unión.

## 5.4) Modificaciones en el sitio 3:

### 5.4.1) Modificaciones en el tamaño del sitio:

Con el objetivo de definir con más precisión la permisividad que presenta la proteína p4 en relación a la longitud de su sitio de unión, se llevaron a cabo ensayos en los que se disminuyó 1 par de bases o se aumentó en 1, 2 ó 3 pares de bases el tamaño del sitio 3. Para modificar lo mínimo posible la secuencia del sitio 3 las inserciones fueron pares de bases A/T y se hicieron entre las posiciones -1 y 0 (figura 36); por la misma razón, la delección de un par de bases del sitio 3 se hizo en la posición 0 (figura 36). Estos DNAs cuya longitud fue modificada se marcaron radiactivamente y se incubaron con diferentes concentraciones de proteína p4. Como se puede ver en la figura 36 la proteína p4 fue capaz de unirse a un sitio 3 modificado con un par de bases más que en sus condiciones naturales, +1T, aunque con una afinidad 5 veces menor. Cuando se aumenta en dos y tres pares de bases el tamaño del sitio 3, la formación de complejos es muy baja para +2T (50 veces más baja) y negativa para +3T. Además del tamaño del sitio, hay que tener en cuenta que con la adición de cada par de bases se está introduciendo un giro de 34° en el DNA. Esto implica que la posición de las guaninas dentro del surco mayor se modifica y probablemente afecte al correcto establecimiento de la interacción específica que se produce con las Arg6 del Garfio de la proteína. Cuando se disminuyó el tamaño del sitio en un par de bases, el efecto en la interacción con la proteína fue mucho más drástico, puesto que se redujo la unión entre 10 y 30 veces con respecto a la unión en la secuencia salvaje, mientras que con el aumento de un par de bases la disminución solo fue de 5 veces.

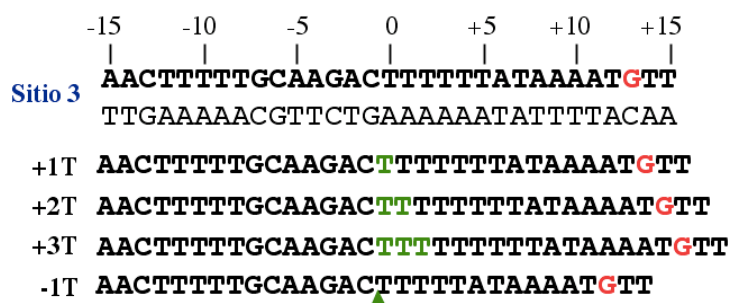
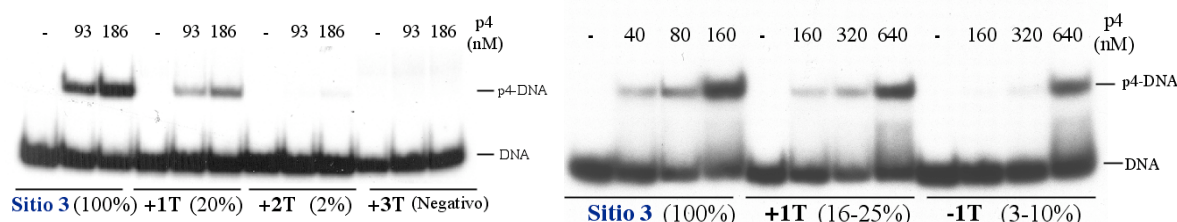


Figura 36: Estudio de la influencia de la longitud del sitio 3 en la interacción con p4. Los nucleótidos insertados y el lugar de la delección en la secuencia se muestran en verde. Se presenta, también, el porcentaje de complejos formados con la proteína p4, con respecto a los complejos formados con el sitio 3 salvaje (100%). Los fragmentos de DNA se marcaron radiactivamente, se incubaron con las concentraciones señaladas de p4 y se resolvieron los complejos p4-DNA mediante electroforesis a 4°C en geles nativos de poliacrilamida al 6%.



La interacción que tiene lugar entre el dímero de p4 y la región central del sitio 3 se produce a través de los residuos Lys51 y Arg54 localizados en la región de dimerización. Para que estas interacciones con los fosfatos de los pares de bases A/T del centro del sitio puedan tener lugar de forma correcta es posible que sea necesario que el DNA presente una determinada holgura, la cual no existiría en un DNA de tamaño menor. Con la intención de analizar si esto era así, se realizaron ensayos en los que se

comparó la afinidad de las proteínas K51A y R54A por los sitios 3, +1T y -1T. En el apartado 4.2 de este trabajo (figura 24) se analizó la unión de ambas proteínas mutantes al sitio 3 sin modificar y se vio que era igual que la de la proteína silvestre. Como se puede ver en la figura 37, cualquiera de las dos modificaciones en el DNA disminuye la capacidad de unión de las proteínas. Cuando se aumenta un par de bases el tamaño del sitio se puede ver que la presencia de las mutaciones afecta a la interacción, reduciéndose tres veces aproximadamente la formación de complejos. Cuando se acorta el tamaño del DNA, el efecto es mucho más drástico, no detectándose la formación del complejo a las concentraciones de proteína analizadas (figura 37).

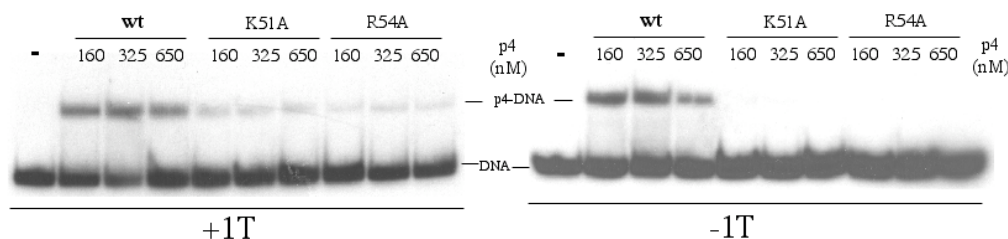


Figura 37: Interacción de las proteínas p4 salvaje, K51A y R54A con los DNAs del sitio 3 +1T y -1T. Tras la incubación con diferentes concentraciones de las proteínas se llevó a cabo electroforesis a 4°C en geles al 6% de poliacrilamida.

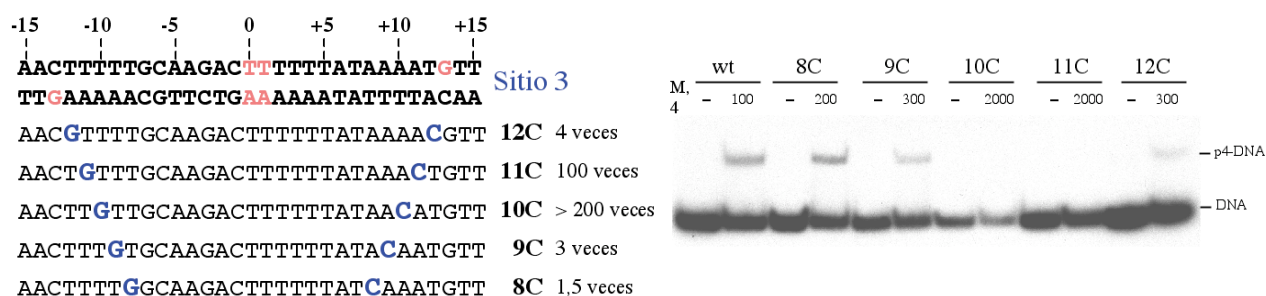
#### 5.4.2) Modificaciones en la secuencia del sitio 3 entre las posiciones 8-12:

Además de la limitación de la longitud de un sitio de unión de p4, con las modificaciones realizadas en las secuencias de los sitios 1, 2 y 4 se definieron dos factores importantes que influyen en la afinidad de unión de p4; la presencia de las guaninas en las posiciones 13 y de pares de bases A/T en las posiciones centrales del sitio (0 y +1). Se decidió, entonces, estudiar la influencia de cada par de bases que forman las series de adeninas y timinas consecutivas de la repetición invertida en la secuencia. Estos grupos de pares de bases A/T tienen estrechado el surco menor en el cristal del complejo p4-DNA. Además, 4 de los aminoácidos modificados en la proteína p4 que más afectaban en la unión al DNA (T4, H10, Y33 y K76) establecen su interacción con el esqueleto fosfato de estas bases A/T.

Se cambiaron los pares de bases A/T por pares de bases C/G, en ambos lados del sitio 3 a la vez, desde las posiciones 12 a las 8 (figura 38A). La formación de los complejos de p4 con estos fragmentos de DNA modificados mostró que, el cambio en las posiciones 8 (8C), apenas reduce la capacidad de unión de la proteína p4 al sitio 3, mientras que cuando se modificaron las posiciones 9 y 12 (9C y 12C) disminuye la formación de los complejos p4-DNA entre 3 y 4 veces. Lo más sorprendente fue el resultado obtenido tras realizar el cambio de A/T por G/C en las posiciones 11 y 10 (11C y 10C), ya que podemos decir que prácticamente no se forma el complejo de la proteína p4 con estos DNAs modificados (100 y más de 200 veces menos, respectivamente)(figura 38A). Teniendo en cuenta el efecto tan drástico que produce la modificación de estas bases, realizamos otro tipo de cambio en las mismas posiciones que consistió en introducir timinas en lugar de citosinas (figura 38B). El resultado mostró que la p4 sí es capaz de interaccionar con estos nuevos fragmentos modificados en las posiciones 10 y 11 (10T y 11T), aunque la formación de los complejos es menor que con la secuencia del sitio 3 silvestre. Estos resultados nos revelaron la importancia de dos nuevas posiciones en el sitio 3 y nos permitió perfilar un poco más las

características más significativas de un sitio de unión de p4 y especular sobre cuáles podrían ser las razones por las que la proteína posee más afinidad por unos sitios de unión que por otros.

A



B

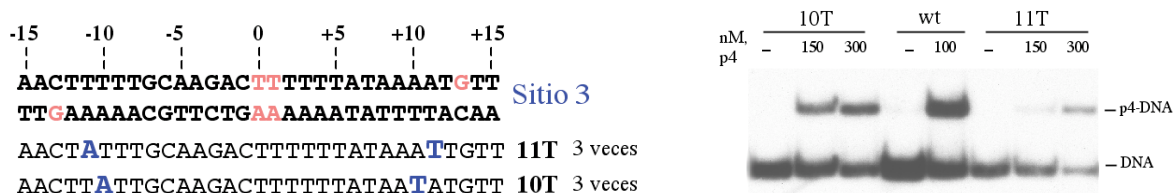


Figura 38: Ensayo de interacción de p4 con el DNA del sitio 3 modificado. Cada uno de los fragmentos de DNA tiene modificadas dos bases, que están en la misma posición respecto al centro (0), pero una en cada lado de la secuencia de DNA (en azul). Los nucleótidos en rojo muestran las dos características hasta ahora demostradas más importantes en la interacción de p4 con el sitio 3. La disminución en la formación de complejos se presenta al lado de cada secuencia como número de veces que se une peor p4 que al sitio 3 salvaje. Los complejos se resolvieron por electroforesis en geles al 6% de poliacrilamida y a 4°C.

Pensamos que el hecho de que, en estas dos posiciones, el cambio a G/C afectase tanto a la unión de p4 podría deberse a la ruptura de la serie de pares de bases A/T consecutivas que forman las repeticiones invertidas del sitio. Por otra parte, el origen del defecto en la interacción podría estar en que un par de bases C/G es menos deformable que un par de bases A/T, debido a la presencia de un puente de hidrógeno más y de un grupo amino en el surco menor con un volumen que opondría resistencia a un posible estrechamiento del mismo. Puesto que la posición 10 parecía la más afectada, quisimos estudiar esta posibilidad y para ello utilizamos análogos de nucleótidos (figura 39). El nucleótido 2-diaminopurina (DAP) es una purina que se aparea con la timina, de modo que se mantendría el grupo de pares de bases A/T, pero aportaría un enlace más y un grupo amino en el surco menor al igual que la guanina. Por otro lado, empleamos la inosina (I) que se aparea con la citosina y, aunque se rompe la serie de adeninas y timinas, se establecen solo dos enlaces y no presenta el grupo amino de la guanina. Se modificó la posición 10 del sitio 3 con estos dos análogos de nucleótido y se analizó la interacción de p4 con los sitios 3 modificados (figura 39). El resultado mostró que p4 era incapaz de interaccionar con el sitio 3 que poseía el nucleótido DAP mientras que con el que poseía la inosina formaba complejos estables, aunque 6 veces menos que con el DNA silvestre.



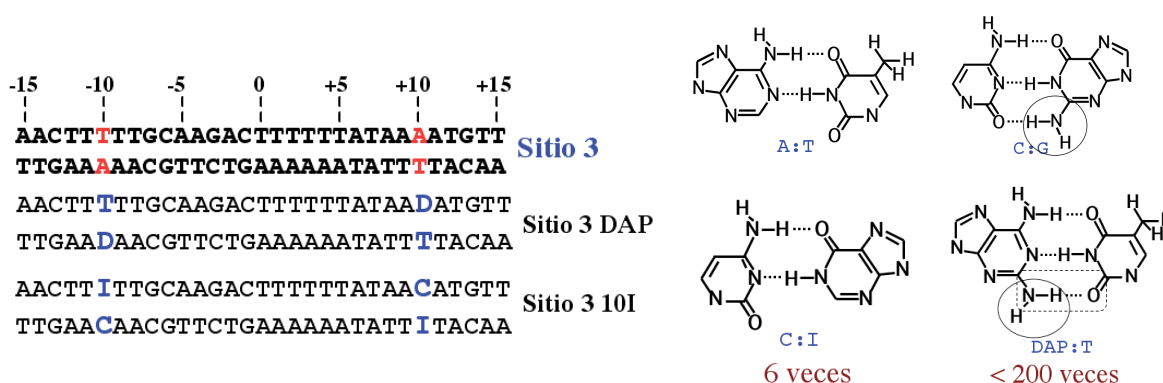


Figura 39: Interacción de p4 con sitios 3 modificados con análogos de nucleótidos. Se presenta la secuencia de los DNAs utilizados en el ensayo y la estructura de los pares de bases de Watson y Crick (G/C y A/T) y los enlaces con los nucleótidos modificados; en círculos se muestra el grupo amino de la guanina y del DAP y en el cuadro discontinuo el tercer enlace del apareamiento DAP:T. (Inosina, I; 2-diaminopurina, DAP).

#### 5.4.3) Diferencias en los extremos del sitio 3:

El sitio 3 no posee una simetría perfecta en sus extremos. El extremo derecho tiene la secuencia 5'AAAAGT 3' y el izquierdo la secuencia 5'AAAATGTT 3'; en la posición +12 se rompe el grupo de 5 adeninas consecutivas, presentes en el extremo izquierdo, por la presencia de una timina. Quisimos estudiar si esta diferencia en la secuencia de los extremos tenía alguna implicación en la afinidad que presentaba la proteína p4 por el sitio 3. Para ello, diseñamos dos fragmentos de DNA con la secuencia del sitio 3 pero que tenían modificados sus extremos. En un caso, ambos extremos eran iguales al izquierdo del sitio; tenían por tanto, grupos de 5 adeninas consecutivas (5'AAAAGT 3') y se le llamó **Sitio 3A**, mientras que en el otro fragmento diseñado, ambos extremos poseían la interrupción del grupo de adeninas en las posiciones 12 (5'AAAATGTT) y se le llamó **Sitio 3B** (figura 40). El resultado de la interacción de la proteína p4 con estos sitios 3 modificados mostró que cuando ambos extremos poseían el grupo de adeninas interrumpido en la posición 12 por una timina, la unión de la proteína p4 al DNA era 4 veces mejor que cuando el sitio 3 presentaba en sus dos extremos el grupo de 5 adeninas consecutivas. Esto indica, por lo tanto, que la asimetría de los extremos del sitio 3 afecta a la afinidad de la proteína. Denominamos “monómero A” al monómero de p4 que interacciona con la parte de la secuencia que posee las 5 adeninas consecutivas y cuya interacción es la más débil, y “monómero B” al que lo hace con la secuencia interrumpida en posición 12 (5'AAAATGTT 3') y que establece la interacción más fuerte con el DNA.

Conociendo la importancia que tienen las guaninas en posición 13 y la asimetría de los extremos del sitio 3, analizamos cómo afectaba a la formación de complejos con p4 la modificación de las guaninas en los extremos del sitio a adeninas. Los fragmentos de DNA diseñados y el resultado del ensayo de interacción se puede observar en la figura 40. En el caso de que la guanina modificada fuese la del extremo derecho, es decir, en posición +13 (**Sitio 3GB**) la unión de la proteína p4 al DNA era aproximadamente 30 veces peor que al DNA silvestre, mientras que cuando el cambio se producía en la posición -13 (**Sitio 3GA**), la unión era aproximadamente 10 veces menor que la salvaje, claramente mejor que la anterior. Estos resultados refuerzan los

datos que indicaban que ambos extremos del sitio no eran iguales, y que p4 tenía mayor afinidad por el extremo donde se unía el monómero B.

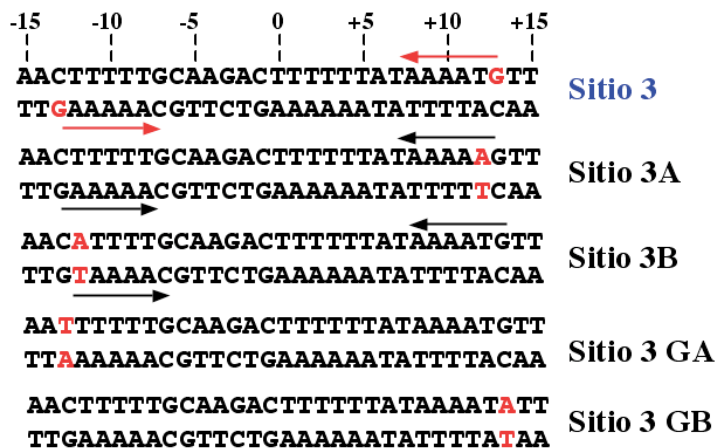
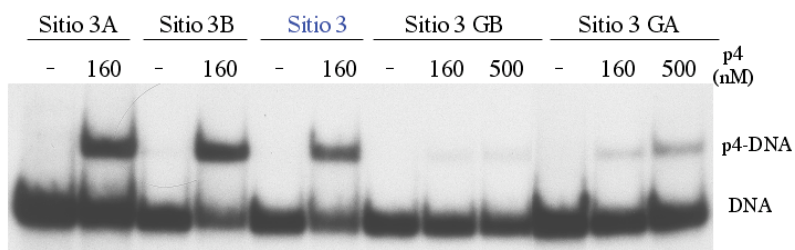


Figura 40: Análisis de la implicación de la asimetría de la secuencia de los extremos del sitio 3 en la afinidad de unión de p4. Se presenta la secuencia de los fragmentos analizados y la autorradiografía del ensayo comparativo de la formación de los complejos, en electroforesis a 4°C en geles nativos de poliacrilamida al 6%



## **6.- ESTUDIO DE LA IMPORTANCIA DEL NÚMERO DE SITIOS Y DE LA SECUENCIA EN LOS EXTREMOS DE CADA SITIO**

Como se ha comentado anteriormente, la proteína p4 forma parte del complejo nucleoproteico que regula la expresión de los genes de  $\phi 29$ . La función de este complejo consiste en reprimir la transcripción del fago que tiene lugar desde los promotores tempranos A2c y A2b y simultáneamente activar la transcripción tardía desde el promotor A3. La proteína p4 tiene cuatro sitios de unión localizados en la secuencia de 219 pares de bases que separa los promotores A2c y A3. La proteína p4 se une como dímero y presenta diferente afinidad, tanto por los distintos sitios de unión, como por cada extremo de un mismo sitio. Teniendo todo esto en cuenta y el hecho de que las guaninas localizadas en posición 13 sean uno de los determinantes más importantes para la unión de p4 a sus sitios, nos propusimos estudiar cuál era el papel de cada sitio en la regulación global de la transcripción, y la implicación de los extremos de cada sitio en la formación del complejo funcional de transcripción. Todo esto en el contexto de los cuatro sitios, los tres promotores y en presencia de todas las proteínas implicadas en el sistema de regulación de la transcripción del fago. Para estudiarlo, se llevó a cabo, mediante mutagénesis dirigida, el cambio de forma individual de cada una de las guaninas en posición  $\pm 13$  de los 4 sitios a citosinas, salvo en el caso del sitio 4, que se modificó por citosina la única guanina que tiene y se insertó una guanina en el extremo en el que carece de ella. Las mutaciones y la nomenclatura final de cada DNA mutante se recogen en la figura 41A.

En primer lugar quisimos analizar la capacidad de la proteína p4 para interaccionar con las secuencias modificadas. Este análisis lo llevamos a cabo estableciendo dos grupos de mutantes; por un lado, analizamos la unión de la proteína p4 a los mutantes de la región 1 que incluye los sitios 1 y 2 y por otro, aquellos con modificaciones en los sitios 3 ó 4 que forman la región 2.

La formación de complejos de la proteína p4 con los fragmentos de DNA modificados de la región 1 se puede observar en la figura 41B. Las mutaciones que implican la pérdida de alguna de las guaninas del sitio 1 (G11 y G12) producen una disminución en la formación de complejos, entre 2 y 10 veces menos. En el caso de las mutaciones en el sitio 2, cuando la guanina modificada fue la correspondiente a la posición +13 (G22), el comportamiento de la proteína p4 es muy similar al que se produce con el DNA sin mutar; en cambio, cuando la mutación es en la guanina -13 (G21), la capacidad de la proteína p4 para formar los complejos se reduce a la mitad.

Los resultados obtenidos con los fragmentos de DNA modificados en la región 2 se pueden ver en la figura 41C. No se observa ningún cambio en la interacción de p4 con el DNA al añadirle al sitio 4 la guanina de la que carece (G4+) y que previamente habíamos visto que mejoraba la capacidad de la proteína p4 para unirse al sitio 4 de forma individual (figura 35). La eliminación de la guanina en posición -13 que posee el sitio 4 (G41) afecta ligeramente a la afinidad de p4 por esta región, disminuyendo entre un 10% y un 20% la formación de complejos. Por otro lado, en el sitio 3, que es por el que p4 tiene mayor afinidad, el resultado fue inesperado, ya que la modificación en la guanina en posición -13 (G31) implica solamente una disminución del 10% en la formación de complejos de p4 con el DNA de la región 2, mientras que la mutación G32 en la posición +13, reduce drásticamente la interacción de p4 con este DNA.



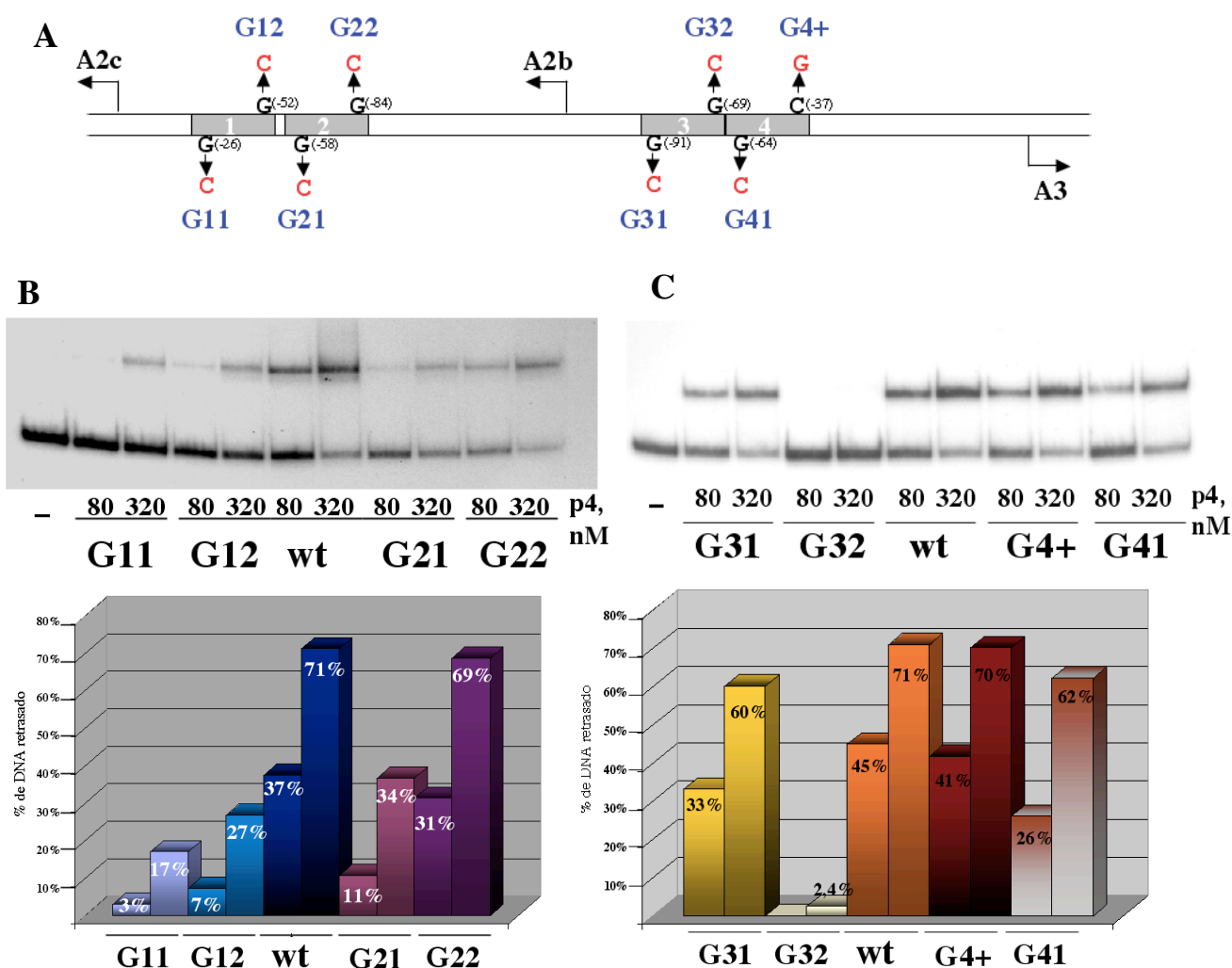


Figura 41: En A, se muestra el esquema de la localización de cada mutación realizada en el genoma de  $\phi 29$ . Se presenta en negro el nucleótido cambiado, en rojo el nuevo nucleótido incorporado y en azul el nombre que se le dio al DNA que lleva dicha modificación. La posición de cada uno de los cambios se numeró desde el promotor A2c en las mutaciones de la región 1 y desde el A3 en las mutaciones de la región 2. En B y C se muestra el resultado de la interacción de la proteína p4 con cada uno de estos DNAs modificados, en la región 1 (B) y en la región 2 (C). Se presenta también con una gráfica de barras el resultado de la cuantificación de la cantidad de los complejos formados.

El conjunto de estos resultados nos indica que en el caso de la región 2 prácticamente toda la fuerza de la interacción de la proteína p4 tiene lugar a través del sitio 3 y, en concreto, por el extremo derecho del sitio. En el caso de la región 1, parece que ambos sitios, el 1 y el 2, están aportando afinidad a la unión final de p4, pero las guaninas del sitio 1 son más críticas para la interacción, puesto que la modificación de cualquiera de ellas reduce mucho la formación de complejos.

Con la intención de profundizar más en la relevancia e implicación de cada uno de los sitios en el sistema de regulación de la transcripción en la región de los promotores A2c, A2b y A3 decidimos desglosar los ensayos para cada uno de los diferentes sitios.

### 6.1) Sitio 1:

Como se ha visto en los experimentos anteriores, la modificación de las guaninas en posición 13 del sitio 1 disminuye la cantidad de complejos que p4 forma con el DNA de la región 1. Por lo tanto, decidimos analizar si su capacidad para formar complejos con las otras proteínas del sistema de regulación, la p6 y la RNAP, estaba o no afectada. Para ello, llevamos a cabo ensayos de protección del DNA a la digestión con DNasa I que permiten estudiar si se forman o no los complejos y cómo son dichos complejos. Los ensayos se realizaron con fragmentos de DNA de las dos secuencias mutadas, G11 y G12, y del DNA silvestre, que incluían desde la posición +41 del promotor A2c hasta la +28 del A3. Son, por lo tanto, fragmentos de DNA que contienen los cuatro sitios de unión de p4 y los tres promotores, A2c, A2b y A3. Se marcaron radiativamente y se incubaron con las diferentes proteínas a las concentraciones indicadas (figura 42). La presencia de la proteína p4 unida en los sitios 1 y 2 es poco evidente (carriles d). En presencia de la RNAP (carriles c) el patrón de bandas es muy claro; se puede ver la huella de la proteína p4 en el sitio 2 y la unión de la RNAP en el promotor A2c. No se aprecian diferencias en la formación del complejo p4-RNAP en ninguno de los DNAs mutados en el sitio 1 con respecto al DNA salvaje. Cuando se analiza la formación del complejo p4-p6-DNA (carriles e) se puede observar que el patrón de bandas del complejo es igual para los 3 DNAs desde el sitio 4 hasta aproximadamente la posición -60 del promotor A2c. A partir de esta posición, muy próxima al sitio 1, existen diferencias en las huellas de los tres DNAs analizados. En el DNA salvaje se protege la

secuencia desde la posición -34 hasta la -10, en G11 esta huella no se produce (líneas negras). En el caso del DNA G12, el patrón de bandas está modificado con respecto al del DNA salvaje, entre la posiciones -60 y -34 (líneas azules); por debajo de la posición -34 se recupera la pauta del DNA salvaje.

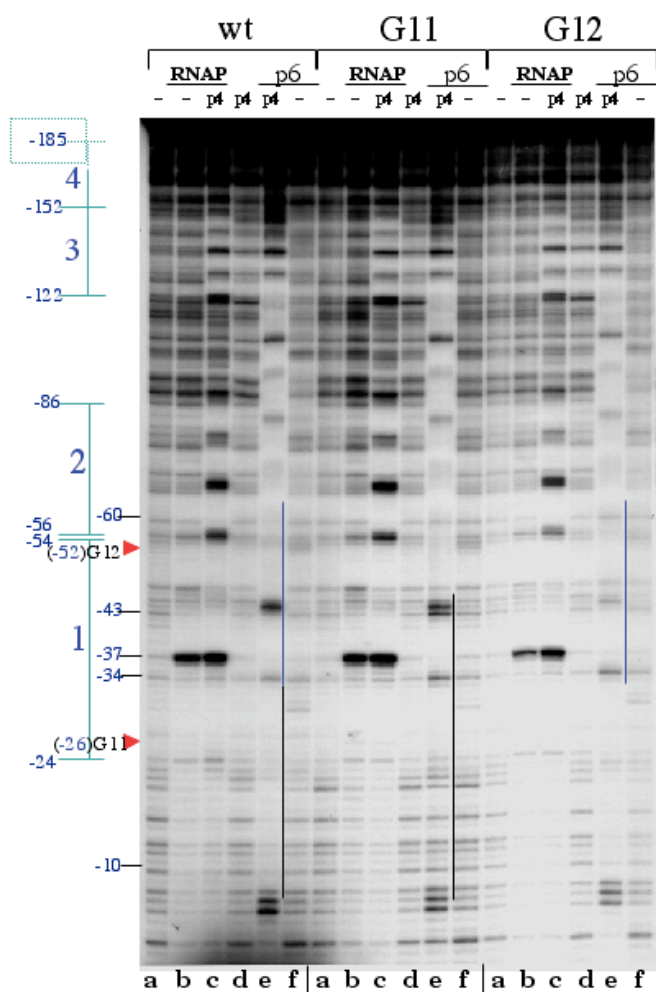


Figura 42: Formación de los complejos de la proteína p4 con las proteínas p6 y RNAP en los DNAs G11 y G12. Los fragmentos de DNA de 288 pb se marcaron radiativamente en el extremo 5' de la posición +41 del A2c. Se incubaron con las proteínas RNAP (94 nM) ó p6 (20  $\mu$ M) y/o p4 (975 nM) y se trataron con la enzima DNasa I. Los fragmentos obtenidos se resolvieron en un gel desnaturalizante al 6% de poliacrilamida-8M urea. Las posiciones que se señalan en la figura están numeradas desde el promotor A2c. Con triángulos rojos se destacan las posiciones mutagenizadas en los DNAs G11 y G12. Las líneas negras muestran las diferencias entre el complejo p4-p6-G11 y el DNA salvaje mientras que con líneas azules se indican las diferencias con el DNA G12. En líneas discontinuas se marca la localización aproximada del extremo del sitio 4.

Estos resultados indican que el defecto que existe en la unión de p4 al DNA tras la mutagénesis en el sitio 1 está afectando a la correcta formación del complejo p4-p6.

Para estudiar la relevancia funcional de estos cambios en la formación de complejos, se llevaron a cabo ensayos de transcripción truncada *in vitro*. Los fragmentos de DNA utilizados fueron los mismos que para el ensayo de DNasa I. El tamaño del transcrito esperado desde el promotor A2b es de 137 nucleótidos, desde el A2c de 41 y desde el A3 de 28 nucleótidos.

Como se puede ver en la figura 43, en presencia de la proteína p4 se produce, en los tres DNAs (G11, G12 y salvaje), una ligera represión del promotor temprano A2c, probablemente debida a que la proteína p4 se une al sitio 2 e interacciona con la RNAP unida al promotor A2c, impidiendo su escape del mismo. Se produce, además, la activación del promotor tardío A3 como resultado del reclutamiento de la RNAP en el promotor, mediado por la interacción con la proteína p4 unida en el sitio 3. Como consecuencia de esta unión, tiene lugar la represión parcial del promotor temprano A2b, cuya caja -35 está en el sitio 3. Cuando se añade al sistema la proteína p6, en el DNA no mutado se observa una fuerte represión de los promotores tempranos A2c y A2b y un aumento en la activación del promotor tardío A3, mientras que en los DNAs G11 y G12 no se observa este efecto; el patrón de transcripción con la proteína p4, es prácticamente el mismo, tanto en presencia como en ausencia de p6. De estos resultados se puede concluir que la función de la unión de la proteína p4 al sitio 1 es requerida para la formación de un complejo nucleoproteico y funcional con la proteína p6 que es el responsable, en definitiva, de la represión de los promotores tempranos A2c y A2b y del aumento de la actividad del promotor tardío A3.

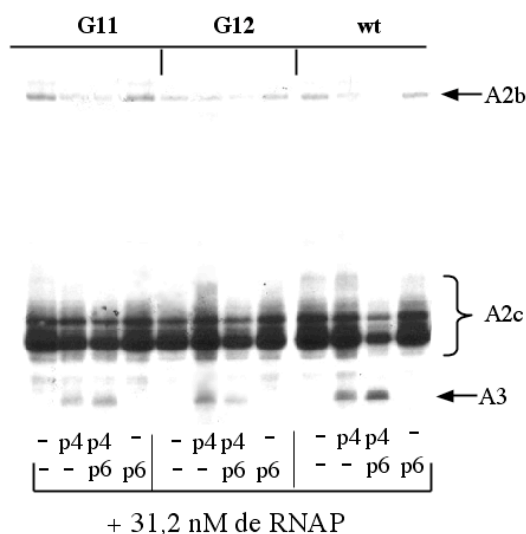


Figura 43: Ensayo de transcripción truncada "in vitro" con fragmentos de DNA modificados en el sitio 1. El DNA posee 288 pb que incluyen los cuatro sitios de unión de p4 y los promotores A2c, A2b y A3 de 29. Tras la incubación de estos DNAs con la RNAP y las proteínas p4 (400 nM) y/o p6 (7  $\mu$ M) se inició la reacción de transcripción añadiendo los cuatro ribonucleótidos, estando el UTP marcado radiativamente en su fósforo  $\alpha$ . La resolución de los transcritos producidos a partir de cada DNA por la RNAP se hizo en geles desnaturalizantes al 6% de poliacrilamida-8M urea.

## 6.2) Sitio 2:

Para estudiar la importancia y la función del sitio 2 llevamos a cabo ensayos de protección del DNA a la digestión con DNasa I y analizamos la formación de los complejos nucleoproteicos que dirigen la regulación de la transcripción en la región central del genoma. Los fragmentos de DNA de 288 pb, con las secuencias G21, G22 y salvaje, se marcaron radiativamente y se incubaron con las proteínas indicadas en la figura 44. El patrón de bandas de los complejos del DNA G22, es idéntico al patrón que muestra el DNA silvestre, de modo que la formación de los complejos con la RNAP y con p6 no se ve afectada por la modificación de la guanina en posición +13 del sitio 2. En cambio, el patrón que produce el complejo p4-RNAP en el mutante G21 es distinto

al que se genera en el DNA silvestre; se pueden ver las diferencias desde el sitio 2 hasta el final del fragmento (líneas negras, carriles c, figura 44), sugiriendo que la interacción p4-RNAP es distinta. La formación del complejo de la proteína p4 con la p6 (carriles e) produjo el mismo patrón en los tres fragmentos de DNA, indicando que las mutaciones en el sitio 2 no afectan a la formación de este complejo.

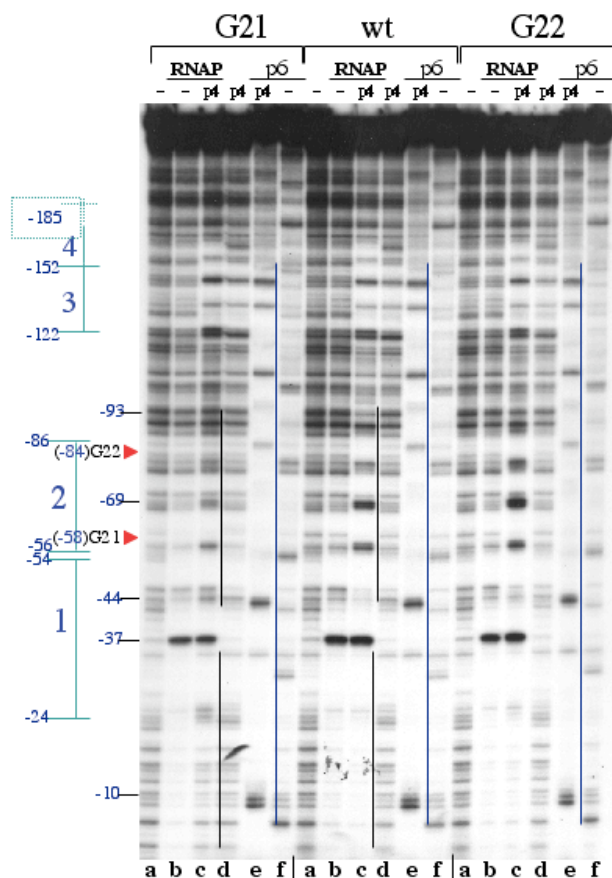


Figura 44: Formación de los complejos de la proteína p4 con las proteínas p6 y RNAP en los DNAs G21 y G22. Los fragmentos de DNA de 288 pb se marcaron radiactivamente en el extremo 5' de la posición +41 del A2c. Se incubaron con las proteínas RNAP (75 nM), p6 (20  $\mu$ M) y/o p4 (975 nM) y se trataron con la enzima DNasa I. Los fragmentos obtenidos se resolvieron en un gel desnaturalizante al 6% de poliacrilamida-8M urea. Las posiciones que se señalan en la figura están numeradas desde el promotor A2c. Con triángulos rojos se destacan las posiciones mutagenizadas en los DNAs G21 y G22. Las líneas negras muestran la región del patrón donde el complejo de p4 con la RNAP es diferente entre el DNA salvaje y el DNA G21, mientras que las líneas azules muestran los complejos p4-p6-DNA que son iguales en los 3 DNAs.

Se estudió la funcionalidad del sistema con estos DNAs mutados en el sitio 2, G21 y G22, por medio de ensayos de transcripción truncada “in vitro”. Como se puede ver en la figura 45, la transcripción con el DNA mutante G22 posee un comportamiento igual al del sistema con el DNA sin mutar; en cambio, con el DNA G21, se observa que la proteína p4 sola no es capaz de reprimir el promotor temprano A2c, lo cual se corrige en presencia de la proteína p6.

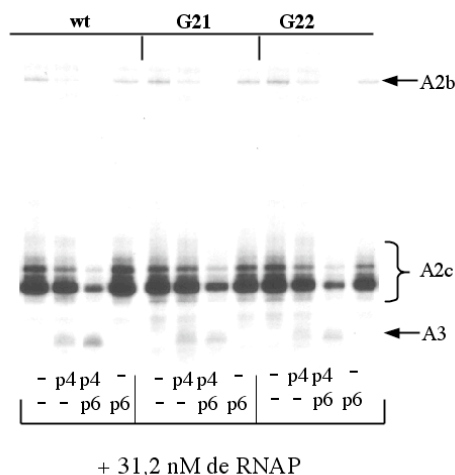


Figura 45: Ensayo de transcripción “in vitro” en fragmentos de DNA con el sitio 2 modificado. El DNA posee 288 pb que incluyen los cuatro sitios de unión de p4 y los promotores A2c, A2b y A3 de  $\phi$ 29. Tras la incubación de estos DNAs con la RNAP y las proteínas p4 (400 nM) y/o p6 (7  $\mu$ M) se inició la reacción de transcripción añadiendo los cuatro ribonucleótidos, estando el UTP marcado radiactivamente en el fósforo  $\alpha$ . La resolución de los transcritos producidos a partir de cada DNA por la RNAP se hizo en geles desnaturalizantes al 6% de poliacrilamida-8M urea

### 6.3) Sitio 3:

La interacción de la proteína p4 con el DNA de la región 2, como vimos anteriormente (figura 41), está ligeramente afectada cuando el DNA tiene mutada la guanina en posición -13 (G31) y muy fuertemente cuando la mutación es en la posición

+13, G32. Quisimos estudiar con detalle la formación de los complejos con las proteínas p4 y p6 en estos DNAs modificados. Se llevaron a cabo ensayos de protección del DNA a la digestión con DNasa I; se utilizaron los mismos fragmentos que en los casos anteriores, pero se marcaron radiativamente en el extremo 5' de la cadena complementaria. Como se puede ver en la figura 46, el patrón de bandas de la unión de la proteína p4 al sitio 3 en los DNAs modificados, es diferente al del DNA silvestre (líneas verdes, carriles b).

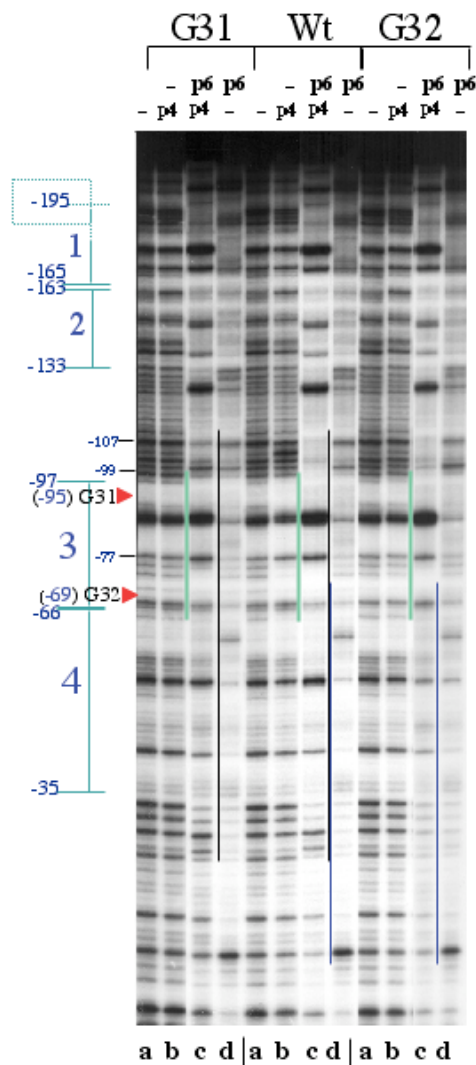


Figura 46: Formación de los complejos de la proteína p4 (1,2  $\mu$ M) con la proteína p6 (20  $\mu$ M) en los DNAs modificados G31 y G32. Los fragmentos de DNA de 288 pb se marcaron radiativamente en el extremo 5' de la posición +29 del A3. Se incubaron con las proteínas indicadas, se trataron con la enzima DNasa I y los fragmentos obtenidos se resolvieron en geles desnaturalizantes al 6% de poliacrilamida-8M urea. Las posiciones que se señalan en la figura están numeradas desde el promotor A3. Con triángulos rojos se destacan las posiciones mutagenizadas en los DNAs G31 y G32. Las líneas negras muestran la región del patrón donde el complejo de la p4 con p6 es diferente entre el DNA salvaje y el DNA G31, mientras que las líneas azules muestran la zona donde se diferencian el DNA salvaje del DNA G32. Las líneas verdes marcan la secuencia del sitio 3, donde no se aprecia unión de la p4 en los DNAs modificados. En líneas discontinuas se marca la localización aproximada del sitio 1.

El complejo de la proteína p4 con la proteína p6 en los DNAs modificados tampoco es igual que en el DNA silvestre (carriles c). En el DNA G31, el patrón de bandas del complejo p4-p6 es similar desde el sitio 1 hasta la zona próxima donde está la mutación (-95), donde se puede destacar la pérdida de protección entre las posiciones -98 y -107 (líneas negras). En el caso de la mutación G32, el complejo es como el silvestre desde el sitio 1 hasta la zona próxima donde está la mutación (-69), y a partir de ahí es distinto del silvestre (líneas azules).

Por lo tanto, con las mutaciones del sitio 3 se ha perdido prácticamente la capacidad de formar correctamente los complejos p4-p6 alrededor de la región 2. Por ello, la regulación de la transcripción debería estar afectada y para comprobarlo realizamos ensayos de transcripción truncada *in vitro* usando como molde los mismos fragmentos de DNA que para los ensayos anteriores. Se analizó la regulación de la transcripción con estos DNAs modificados en el sitio 3 (figura 47). La represión del promotor A2c mediada por el complejo p4-p6 es menor que en el DNA silvestre, debido a que este complejo no se forma correctamente. Además, puesto que en estos DNAs modificados, la p4 no se une adecuadamente al sitio 3, la RNAP no puede ser



estabilizada en el promotor A3 y por lo tanto, este promotor no se activa. El defecto en la unión de p4 al sitio 3 es responsable de la falta de represión del promotor A2b, ya que no compiten por la unión de la RNAP al promotor y además, puesto que el complejo p4-p6 no se forma correctamente, la proteína p4, en presencia de p6, tampoco compete con la RNAP.

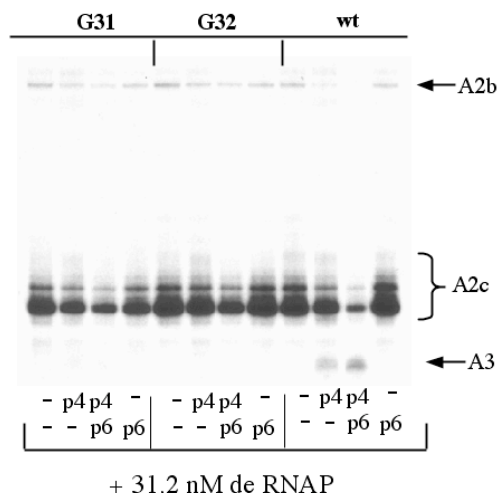
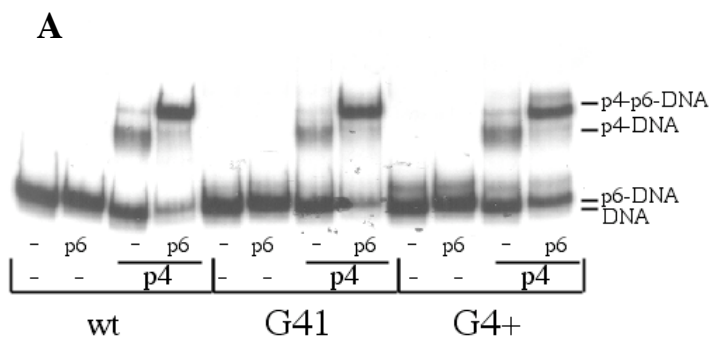


Figura 47: Ensayo de transcripción *in vitro* con fragmentos de DNA con el sitio 3 modificado. El DNA posee 288 pb que incluyen los cuatro sitios de unión de p4 y los promotores A2c, A2b y A3 de  $\phi 29$ . Tras la incubación de estos DNAs con la RNAP y las proteínas p4 (400 nM) y/o p6 (7  $\mu$ M) se inició la reacción de transcripción añadiendo los cuatro ribonucleótidos, estando el UTP marcado radiativamente en su fósforo  $\alpha$ . La resolución de los transcritos producidos a partir de cada DNA por la RNAP se hizo en geles desnaturalizantes al 6% de poliacrilamida-8M urea.

#### 6.4) Sitio 4:

Cuando se analizó la interacción de la p4 con el DNA de la región 2 que poseía modificaciones en el sitio 4 (figura 41) se observó que cuando el cambio consistía en quitar la guanina correspondiente a la posición -13 (G41) disminuía menos de dos veces (15%) la formación de complejos, mientras que cuando se añadía una guanina correspondiente a la posición +13 (G4+) apenas había cambios en la unión a la región 2 con respecto al DNA sin modificar.

Para analizar si las mutaciones en el sitio 4 tenían algún efecto en la formación de complejos de la proteína p4 con la p6, se incubaron los diferentes fragmentos modificados, G41, G4+ y el salvaje, previamente marcados, con dichas proteínas y los complejos formados se resolvieron por electroforesis a 4°C en geles nativos de poliacrilamida. Los fragmentos de DNA utilizados contienen los cuatro sitios y los tres promotores. Como se puede ver en la figura 48A, en ambos se forman los complejos p4-p6 como en el DNA no modificado. Este resultado es coherente con la idea de que la formación del complejo p4-p6 tiene lugar desde el sitio 1 hasta el sitio 3. Dado que la RNAP en el promotor A3 se une en la secuencia del sitio 4 y con la intención de



estudiar si la pérdida o adición de las guaninas podría afectar la formación de los complejos p4-RNAP en este promotor, se llevaron a cabo ensayos de protección a la digestión con DNasa I. El complejo p4-RNAP se forma

en ambos mutantes del mismo modo que en el DNA salvaje (figura 48B).

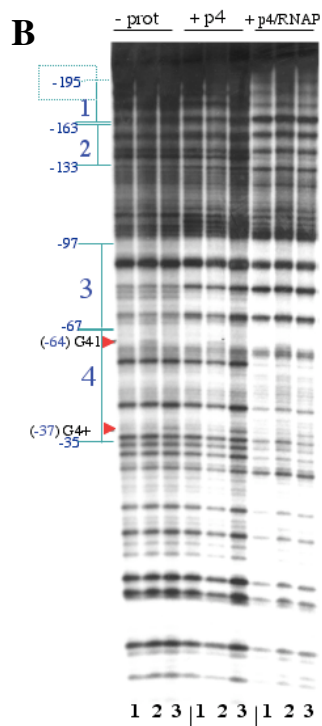


Figura 48: Formación de los complejos en los DNAs G41 y G4+. (A). La formación del complejo de la proteína p4 (160 nM) con la proteína p6 (20  $\mu$ M) se analizó por electroforesis a 4°C en geles al 4% de poliacrilamida. La formación de los complejos de la proteína p4 (800 nM) con la RNAP (75 nM) se estudió mediante ensayos de protección del DNA a la digestión con DNasa I (B); los carriles 1 corresponden al DNA salvaje, los carriles 2 al DNA modificado G41 y los carriles 3 al DNA G4+. Las posiciones señaladas en la figura están numeradas desde el promotor A3; en líneas discontinuas se muestra la localización aproximada del extremo del sitio 1.

Por último, analizamos mediante ensayos de transcripción truncaada *in vitro*, la regulación de la transcripción con estos DNAs mutados. Como se puede ver en la figura 49, las mutaciones no afectan a la regulación de la transcripción, puesto que, con las mismas concentraciones de proteínas, la represión de la transcripción temprana y la activación de la tardía se producen en la misma medida que en el DNA salvaje.

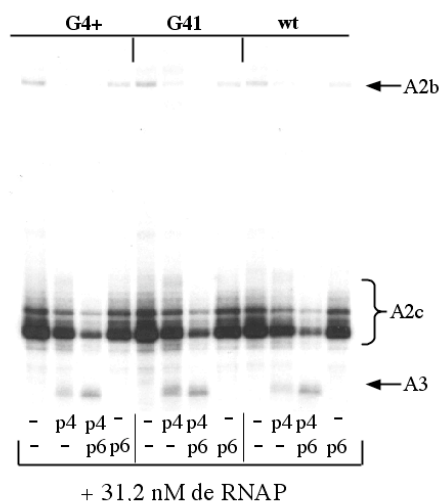


Figura 49: Ensayo de transcripción “in vitro” en fragmentos de DNA con el sitio 4 modificado. El DNA posee 288 pb que incluyen los cuatro sitios de unión de p4 y los promotores A2c, A2b y A3 de  $\phi$ 29. Tras la incubación de estos DNAs con la RNAP y las proteínas p4 (400 nM) y/o p6 (7  $\mu$ M) se dispara la reacción de transcripción con los cuatro ribonucleótidos, estando el UTP marcado radiactivamente en su fosfato  $\alpha$ . La resolución de los transcritos producidos en cada DNA por la RNAP se hizo en geles desnaturalizantes al 6% de poliacrilamida-8M urea.

## 7.- COOPERATIVIDAD EN LA UNIÓN DE p4 A SUS SITIOS

A la vista de estos resultados podemos decir que los cuatro sitios de unión de p4 no tienen ni el mismo papel ni la misma importancia en el proceso de regulación de la transcripción de los genes de  $\phi 29$ . Estos sitios de unión se encuentran organizados por parejas en dos regiones, separadas entre sí por 35 pares de bases. La unión de p4 a cada una de estas regiones genera una curvatura final de aproximadamente  $86^\circ$ . Los sitios 3 y 4 se encuentran colocados uno a continuación del otro, mientras que entre el sitio 1 y el sitio 2 existe un par de bases de separación. Teniendo en cuenta estos datos y el hecho de que en cada sitio se une un dímero de p4, pensamos que la proximidad entre estos sitios contiguos podría implicar la interacción entre dímeros y esto se podría reflejar en la afinidad y eficacia con la que p4 se une al DNA. Nos planteamos, por tanto, estudiar la existencia de cooperatividad en la unión de la p4 a sus sitios. Se usaron cantidades equimolares de fragmentos de DNA marcados radiativamente conteniendo un sitio, dos sitios o los cuatro sitios, y estos fragmentos se incubaron con la misma concentración molar de p4 por sitio, de modo que un fragmento con cuatro sitios se incubó con cuatro veces más cantidad de proteína que uno con un solo sitio de unión. Se cuantificó el porcentaje de complejos de la proteína p4 formados con cada uno de estos fragmentos.

En la figura 50 se recoge en una gráfica el resultado de estos ensayos.

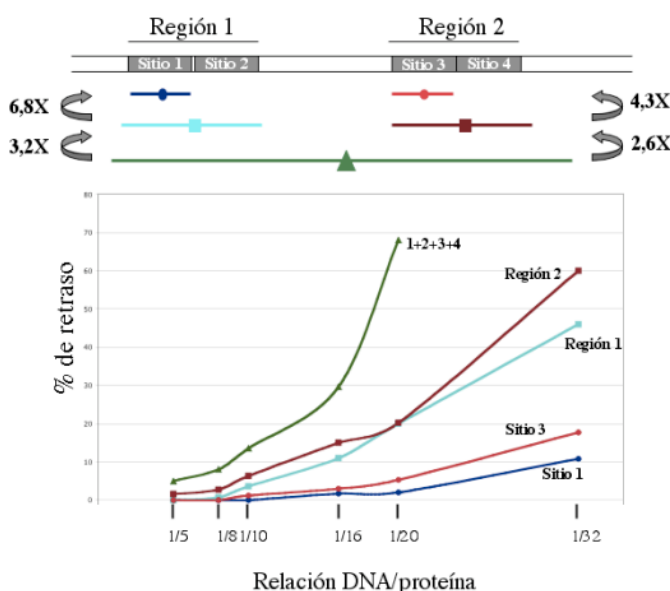


Figura 50: La gráfica muestra la relación que hay entre la formación de complejos de p4 con el DNA en función del número de sitios presentes en el ensayo, a diferentes relaciones molares entre DNA y proteína. En diferentes tonos de azul se presenta el resultado para el sitio 1 y la región 1, mientras que en tonos rojos se representan los resultados obtenidos con el sitio 3 y la región 2. En verde se muestra el resultado con el fragmento de DNA que contiene los cuatro sitios de unión de p4.

La formación de complejos es mayor cuanto mayor es el número de sitios presentes en el ensayo, de modo que la presencia de un segundo sitio mejora entre 4 y 7 veces la formación de complejos de p4 con el DNA, en función de si se trata del sitio 3 en presencia del 4 o del sitio 1 en presencia del 2, respectivamente. Teniendo en cuenta los datos de afinidad entre sitios previamente estudiados, es lógico pensar que el sitio 2 aporte mayor afinidad a la región 1 que el sitio 4 a la región 2. Cuando analizamos lo que sucede entre regiones, se observa que la diferencia es menor. La formación de complejos en presencia de los cuatro sitios es entre 2 y 3 veces mayor que en presencia de sólo 2 sitios. El conjunto de estos datos indica que el aumento en el número de sitios de unión de p4 incrementa el porcentaje de complejos formados y por lo tanto, existe cooperatividad en la unión de p4 a sus sitios, aunque no se trate de una cooperatividad fuerte.



Esta cooperatividad en la unión de p4 en presencia de dos sitios, dada su proximidad, podría deberse a una interacción entre dímeros. En la figura 51A se puede ver en una simulación informática qué regiones de la proteína p4 serían candidatas para mediar la tetramerización de la proteína unida a dos de sus sitios de unión consecutivos. Estas interacciones se podrían dar a través del extremo N-terminal; en concreto, el residuo Gln5 sería un buen candidato puesto que interacciona con la última base de cada sitio, que se encuentran justo a continuación uno del otro (sitios 3 y 4), o separados por una base (sitios 1 y 2). Por otra parte, podría tener lugar a través del extremo C-terminal; la  $\alpha$ -hélice 4 es muy flexible y por ello se seleccionó también como posible mediador de una hipotética tetramerización. Decidimos llevar a cabo ensayos de formación de complejos con el DNA utilizando dos proteínas mutantes, Q5A y  $\Delta 14+2$ ; la primera proteína tiene cambiada la glutamina 5 por alanina y la segunda tiene delecionados los 14 últimos aminoácidos de p4 que implica la pérdida de toda la  $\alpha$ -hélice 4 (Rojo *et al.*, 1990). Se analizó la cooperatividad con estas proteínas modificadas (figura 51B). El resultado mostró que la relación que existe entre la formación de complejos de las proteínas mutadas con el DNA que posee un sitio o dos sitios se mantiene con respecto a la de la p4 silvestre, tanto en la región 1 como en la región 2 (6 y 4 veces, respectivamente). Estos resultados indican que la  $\alpha$ -hélice 4 no es responsable de la posible interacción entre dímeros de p4 y que si la Gln5 está implicada, no lo hace por medio de su cadena lateral.

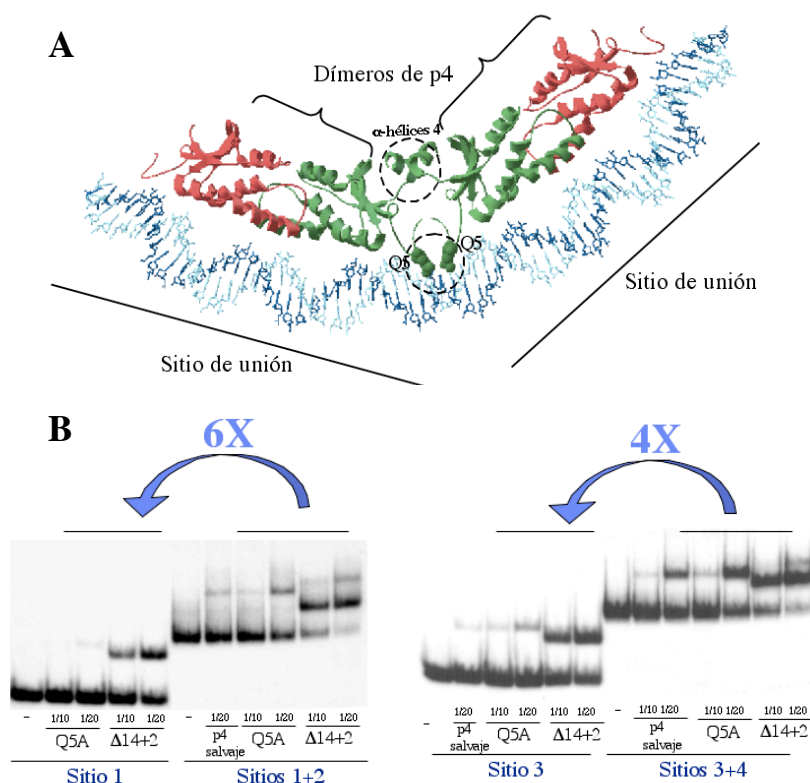
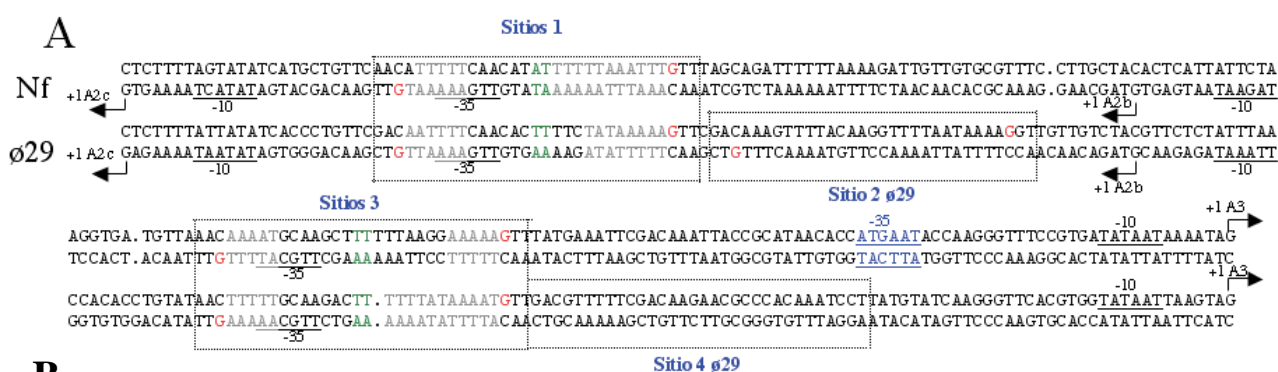


Figura 51: El dibujo (A) es una simulación informática de una posible tetramerización de la proteína p4, partiendo de la estructura de p4 unida al sitio 3. En B se muestran las autorradiografías de un ensayo de formación de complejos con un DNA con la secuencia del sitio 1 o del sitio 3 y con las secuencias de las regiones 1 ó 2, con las proteínas Q5A,  $\Delta 14+2$  y p4 salvaje. La resolución de los complejos se realizó por electroforesis a 4°C en geles nativos de poliacrilamida al 6%. Las dosis empleadas de proteína se presentan como relación de moléculas de DNA/moléculas de proteína. Las flechas azules muestran cuántos complejos más se forman en cada región con respecto a cada sitio.

## 8.- ESTUDIO DE LA REGULACIÓN DE LA TRANSCRIPCIÓN DEL FAGO Nf RELACIONADO CON ø29

La familia de bacteriófagos relacionados con ø29 consta de tres grupos cuya clasificación se realizó atendiendo a diferentes criterios (Pecenkova y Paces, 1999; Yoshikawa 1985; Yoshikawa 1986). Todos ellos tienen un genoma de DNA lineal de doble cadena de aproximadamente 20 Kb de tamaño y poseen una proteína terminal unida covalentemente en sus extremos 5', que es la molécula cebadora en la iniciación del proceso de replicación. El estudio comparativo de los mapas transcripcionales de los fagos ø29 y Nf pertenecientes a los grupos I y II, respectivamente, mostró que eran muy similares (Nuez y Salas, 1993). Nf poseía promotores homólogos a los promotores tempranos A2c y A2b de ø29, ambos con secuencias conservadas en -10 y en -35 y también un promotor homólogo al promotor tardío A3 de ø29, cuyo elemento -35 distaba mucho del hexámero consenso (figura 52) (Nuez y Salas, 1993). La activación de este promotor requería la presencia de una proteína viral, que es idéntica en un 78,6% a la proteína p4 de ø29 (figura 52B) (Mizukami *et al.*, 1986). También se determinó que Nf poseía un sitio de unión para la proteína p4 de ø29; este sitio de unión presentaba homología tanto en secuencia como por su localización con el sitio 3 de ø29 (Nuez y Salas, 1993). Teniendo en cuenta estos antecedentes nos planteamos estudiar el mecanismo de regulación de la transcripción del fago Nf con el objetivo de compararlo con el de ø29 y así poder discriminar cuáles son los aspectos conservados del sistema y por lo tanto más importantes o imprescindibles, y cuáles son aquellos matices que hacen único a cada organismo.



**B**

p4	ø29	MPK <b>T</b> ORGIYH	NLKESEYVAS	NTDVTFFFS	ELYLNKFLDG	YQEYRKFNK
p4	Nf	MPR <b>T</b> ARGIYH	NLKESEYVVS	NGDATFFFS	EMYQNKFLDG	YQKHREFNK
p4	ø29	KIERVAVTPW	NMDMLADITF	YSEVEKRGFH	AWLKGDNATW	REVHVYALRI
p4	Nf	KINRITDTPL	NMDMLADITF	YSNVEKRGFH	AWLKGCNATW	QEIHVYALRT
p4	ø29	MTKPN <b>T</b> LDWS	RIQK <b>P</b> RLRER	RKSMV		
p4	Nf	MTKPCTQ <b>N</b> WS	RIRK <b>P</b> KLVER	RKSMV		
						78,6%
						Identidad

**C**

p6 ø29	.AKMMQREIT	KTVNVAKMV	MVDGEVQVEQ	LPSETFVGNL	TMEQAQWRMK
p6 Nf	MRKMMQREVT	YTTAQLARMK	MVDGEVTAEV	LEPVTLIGNL	SVEQAQREIN
p6 ø29	RK..YKGE <b>P</b> V	QVVSVEPNTE	VYELPVEKFL	EVATVRVEKD	EDQEEQTEAP
p6 Nf	KRSEFENPA	QVVGVEANTQ	LYELPLDVFL	EHATVK.ERP	ATKEEVAEVQ
p6 ø29	EEQVAE				
p6 Nf	A				
					50,5%
					Identidad

Figura 52: Comparación de secuencias de los fagos ø29 y Nf. Se muestra la secuencia de ambos genomas desde sus promotores A2c hasta sus promotores A3. Se han señalado en la secuencia la localización de los sitios de unión de la p4 de Nf determinados por homología, y las características que comparte con los sitios de p4 de ø29. También se señala la secuencia poco conservada del elemento -35 del promotor A3 de Nf. En B y C se muestran las secuencias comparadas de las proteínas p4 y p6 de Nf con las de ø29; en negro los aminoácidos idénticos, en azul los diferentes.

El fago Nf tiene, además, un gen temprano cuya localización en el genoma es homóloga al gen que codifica la proteína p6 de  $\phi 29$ , estas proteínas poseen un 50,5% de identidad en su secuencia aminoacídica (Freire *et al.* 1996)(figura 52C).

Al analizar las secuencias comprendidas entre los promotores A2c y A3 de ambos fagos pudimos observar que, por homología de secuencia, sólo había 2 posibles sitios de unión de la proteína p4, los correspondientes por su posición a los sitios 1 y 3 de  $\phi 29$  (figura 52A). El sitio 3 de Nf tiene un par de bases más, pero en general, las secuencias de los sitios de Nf siguen un patrón muy similar al de los sitios de unión de  $\phi 29$ . Se puede ver que a continuación de las guaninas (en rojo) y hacia dentro del sitio hay grupos de adeninas o timinas consecutivas (en gris) y que los nucleótidos correspondientes a las posiciones 0 y +1 (en verde) están también ocupados en Nf por adeninas o timinas que, como se ha visto anteriormente, son importantes para la afinidad de la p4 de  $\phi 29$  por sus sitios de unión.

Para llevar a cabo el estudio de la transcripción del bacteriófago Nf, en primer lugar se clonó, se secuenció y se purificó la proteína p4 de Nf, a la que denominamos p4N; se detectó que la secuencia de la proteína p4N difería en los aminoácidos de las posiciones 30 (F por S) y 86 (V por C) de la secuencia de esta proteína publicada previamente (Mizukami *et al.*, 1986); confirmamos por secuenciación del gen del genoma de Nf que la p4 purificada era realmente la secuencia de la proteína de la cepa del fago que poseíamos en el laboratorio. Todos los aminoácidos implicados en la interacción p4-DNA en  $\phi 29$  están conservados en la p4 de Nf, salvo la Gln5 (en rojo, figura 52B). Se purificó la proteína p6 de Nf, que se denominó p6N y se clonó también el DNA de la región central de los promotores de Nf.

### 8.1 Interacción de las proteínas p4 de Nf (p4N) y p6 de Nf (p6N) en la región de los promotores A2c, A2b y A3:

Nos planteamos determinar la interacción de las proteínas virales p4N y p6N con el DNA de Nf. Para ello, se llevaron a cabo ensayos de retraso de la movilidad electroforética de un fragmento de DNA que contenía la secuencia comprendida entre la posición +67 del A2c y la +48 del A3 del fago Nf y se incubó con la proteína p4N, con la p6N o con ambas proteínas.

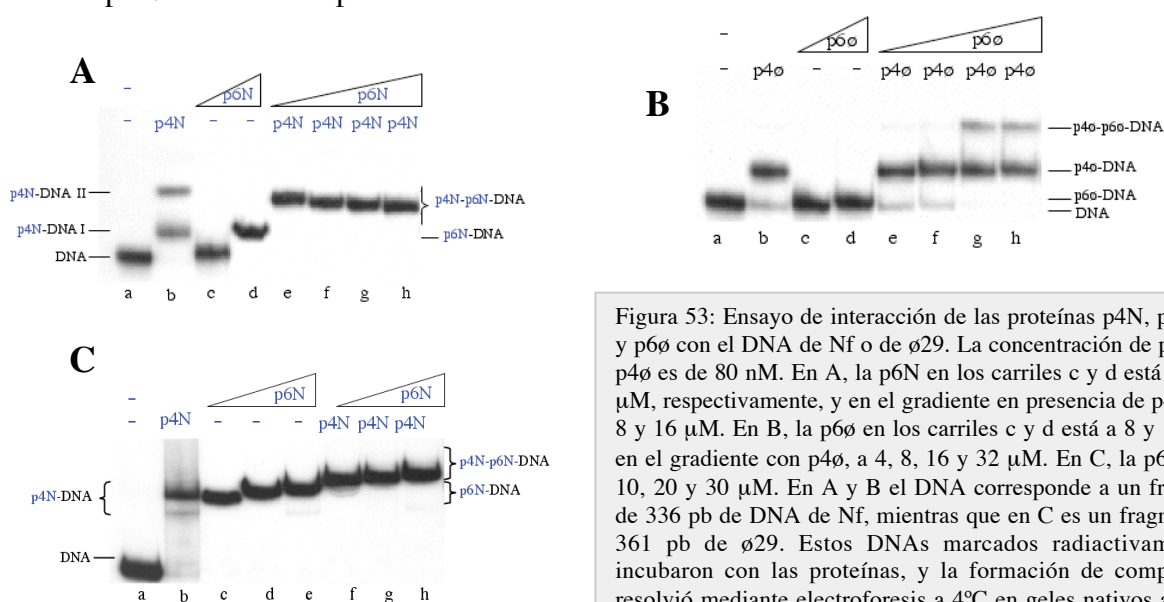


Figura 53: Ensayo de interacción de las proteínas p4N, p4 $\phi$ , p6N y p6 $\phi$  con el DNA de Nf o de  $\phi 29$ . La concentración de p4N y de p4 $\phi$  es de 80 nM. En A, la p6N en los carriles c y d está a 8 y 16  $\mu$ M, respectivamente, y en el gradiente en presencia de p4, a 2, 4, 8 y 16  $\mu$ M. En B, la p6 $\phi$  en los carriles c y d está a 8 y 16  $\mu$ M y en el gradiente con p4 $\phi$ , a 4, 8, 16 y 32  $\mu$ M. En C, la p6N está a 10, 20 y 30  $\mu$ M. En A y B el DNA corresponde a un fragmento de 336 pb de DNA de Nf, mientras que en C es un fragmento de 361 pb de  $\phi 29$ . Estos DNAs marcados radiactivamente se incubaron con las proteínas, y la formación de complejos se resolvió mediante electroforesis a 4°C en geles nativos al 4% de poliacrilamida.

Como se puede ver en la figura 53A, la unión de la proteína p4N al DNA de Nf da lugar a la aparición de dos bandas de movilidad diferente, una de las bandas podría corresponder a la proteína p4 unida en sólo uno de los sitios y la otra al complejo formado por la unión de la proteína a ambos sitios. Cuando se analiza la interacción de la proteína p6 al DNA se observa que a dosis bajas (8  $\mu$ M, carril c) no se aprecia la interacción y que al aumentar la cantidad de p6 al doble se puede ver la formación del complejo. Para estudiar la formación de un complejo de las proteínas p4N y p6N en el DNA de Nf, se incubó una concentración fija de p4N con concentraciones crecientes de p6N. Como se puede ver en la figura 53A (carriles e-h), se obtiene sólo una banda retrasada; a bajas concentraciones de p6 la movilidad de este complejo es menor que la del complejo p6N-DNA y que la de una de las especies de p4N-DNA; a medida que se va aumentando la cantidad de p6, el complejo adquiere una mayor movilidad. Esto sugiere que existe un cambio en la conformación del complejo p4N-p6N-DNA al aumentar la concentración de p6, o bien que existe un cambio en la carga iónica neta del complejo. De este ensayo, además, se puede concluir que en la formación del complejo p4N-p6N con el DNA las proteínas se comportan de forma sinérgica, como ocurre en el sistema de  $\phi$ 29, puesto que a 8  $\mu$ M de p6N no se observa formación del complejo p6-DNA (línea c), mientras que en presencia de p4N, 2 $\mu$ M de p6N es suficiente para obtener el complejo p4N-p6N-DNA.

En paralelo se llevó a cabo un ensayo con las proteínas p4 y p6 de  $\phi$ 29 en el mismo DNA de Nf y en idénticas condiciones. En la figura 53B se puede ver que la proteína p4 de  $\phi$ 29 con el DNA de Nf da lugar sólo a un complejo, lo cual podría significar que, o bien la unión de p4 de  $\phi$ 29 a un sitio o a los dos sitios de Nf da lugar a especies que poseen la misma movilidad electroforética, o bien la proteína p4 solamente es estable unida a uno de los sitios de Nf. La unión de la proteína p6 de  $\phi$ 29 al DNA de Nf da lugar a un complejo de movilidad ligeramente menor al que posee el DNA sin proteínas. Como se puede ver en la figura 53B, en las líneas de e-h, se necesitó ocho veces más cantidad de p6 $\phi$  que de p6N, para formar el complejo p4 $\phi$ -p6 $\phi$ -DNA, y menos del 50% del DNA forma el complejo. Por lo tanto, las proteínas p4 y p6 de  $\phi$ 29 en el DNA de Nf no se comportan de forma sinérgica; de hecho, en estos carriles, se observan tres bandas, la del complejo DNA-p6, la correspondiente a DNA-p4 y por último el complejo DNA-p4-p6.

Decidimos, entonces, analizar la formación de los complejos con las proteínas p4N y p6N en el DNA de  $\phi$ 29. Utilizamos para ello un fragmento de DNA de 361 pb con la secuencia de los promotores A2c, A2b y A3 de  $\phi$ 29 equivalente a la usada del DNA de Nf. Como se puede ver en la figura 53C, la p4N forma un complejo con el DNA de  $\phi$ 29 que retrasa el 100% de las moléculas, al igual que hace la p6N desde la concentración más baja ensayada. En presencia de las dos proteínas, p4N a una concentración fija y concentraciones crecientes de p6N, se forman complejos de menor movilidad, que van disminuyendo al ir aumentando la cantidad de proteína p6N.

Las proteínas p4N y p6N forman, por tanto, en el DNA de los promotores A2c, A2b y A3 de  $\phi$ 29, complejos muy similares a los que forman en su propio DNA, mientras que las proteínas p4 y p6 de  $\phi$ 29 parecen tener menor afinidad de unión y menor capacidad para formar el complejo p4-p6 en la secuencia de DNA de Nf.

### 8.2) Determinación de los sitios de unión de p4N en su genoma:

Nos propusimos averiguar cuántos sitios de unión tenía la proteína p4N en el genoma de Nf; como se ha comentado anteriormente, por homología de secuencia con el DNA de  $\phi 29$ , solamente se encontraron dos posibles sitios (figura 52). Además, tras los ensayos de interacción de p4N con el DNA de Nf, se observó que se formaban dos complejos con movilidades diferentes (figura 53) lo cual, nuevamente podría indicar que eran dos los sitios de unión de la proteína en su genoma. Para analizar definitivamente dónde se unía la proteína p4N en el DNA de Nf, se llevaron a cabo ensayos de protección del DNA a la digestión por DNasa I (figura 54). La unión de la proteína p4N se produce en dos zonas (entre -24 y -54 y entre -120 y -151), y se caracterizan por la aparición de hipersensibilidades y de protecciones en la secuencia (carriles 2, 3 y 4). Estas zonas se corresponden con los sitios 1 y 3 de Nf descritos por homología de secuencia con  $\phi 29$ . El sitio 3 coincide con la secuencia a la que se une la p4 de  $\phi 29$  en el DNA de Nf descrita en Nuez *et al.* (1993). En las posiciones de la secuencia relativas a las que ocuparían los sitios 2 y 4 de  $\phi 29$  en el DNA de Nf no se detecta ninguna modificación del patrón, lo que nos permite confirmar que p4N posee únicamente dos sitios de unión que se han denominado sitio 1 y 3 de Nf.

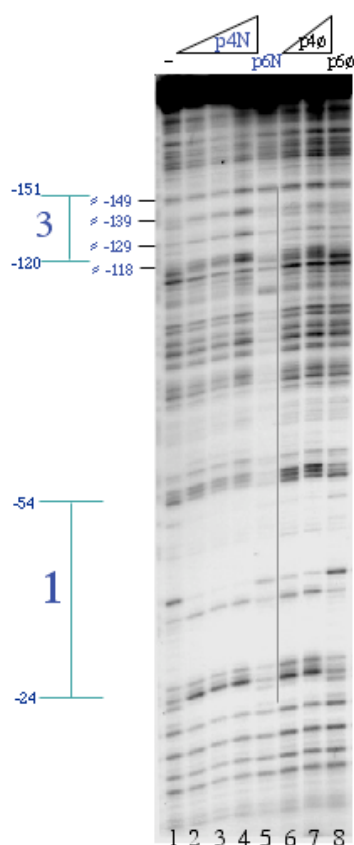


Figura 54: Estudio del número de sitios de unión de p4N en el DNA de Nf. Se marcó radiactivamente el extremo 5' de la posición +24 desde el A2c de un fragmento de 259 pb. Se incubó con las proteínas indicadas. Las concentraciones de p4N que se utilizaron fueron 0,4, 0,8 y 1,6  $\mu$ M, de p4 $\phi$  fueron 0,4 y 1,6  $\mu$ M y de ambas proteínas p6, 20  $\mu$ M. Posteriormente, este DNA se trató con la DNasa I y se resolvieron los fragmentos obtenidos en un gel desnaturalizante al 6% de poliacrilamida-8M urea. Las posiciones están numeradas desde el promotor A2c. La línea gris muestra la unión de p6N.

Se analizó la unión de la proteína p4 de  $\phi 29$  en esta misma secuencia de Nf; como se puede ver en los carriles 6 y 7 (figura 54) la proteína genera un patrón similar al que produce p4N, pero más débil, lo que sugiere que posee menor afinidad por estos sitios de unión que p4N. Por otra parte, se puede observar que el patrón producido en el DNA por la p4 de  $\phi 29$  es más claro sobre la secuencia del sitio 1 que sobre la secuencia del sitio 3, lo cual podría implicar que posee mayor afinidad por el sitio 1, al contrario de lo que ocurre con los sitios 1 y 3 de su genoma.

En los carriles 5 y 8 (figura 54) se muestra la unión de las proteínas p6 de Nf y  $\phi 29$ , respectivamente. La proteína p6 de Nf produce un patrón de bandas e hipersensibilidades que abarca prácticamente toda la secuencia, en concreto, desde la posición -20 a la -150 del A2c (línea gris), mientras que la p6 de  $\phi 29$  a la concentración ensayada (20  $\mu$ M) no se une al DNA de Nf, puesto que se genera un patrón de bandas igual en presencia y en ausencia de la proteína (comparar carriles 1 y 8).

Quisimos profundizar en la interacción de las proteínas p4 con los sitios de unión de Nf, utilizamos un fragmento de DNA que solamente contenía como secuencia específica la del sitio 3 de Nf y analizamos la afinidad de unión de las proteínas p4 por este DNA; aunque en ningún caso se obtuvieron complejos muy estables, se pudo concluir que la p4N se une aproximadamente 5 veces mejor que la p4 de  $\phi 29$  al sitio 3 de Nf (figura 55). También se analizó la unión al sitio 1 de Nf de ambas proteínas p4 y observamos que las dos presentaban una ligera mayor afinidad por el sitio 1 que por el sitio 3 de Nf (no mostrado).

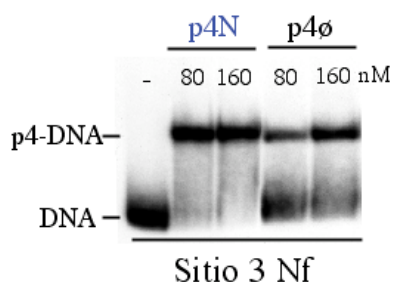


Figura 55: Interacción de las proteínas p4 de Nf y de  $\phi 29$  con el sitio 3 de Nf. El DNA se marcó radiactivamente y se incubó con las concentraciones de proteína indicadas. La formación de complejos se resolvió por electroforesis a 4°C en geles nativos de poliacrilamida al 6%

Puesto que en  $\phi 29$  la formación de un complejo en el DNA con las proteínas p4 y p6 es fundamental para la regulación de la transcripción del sistema y teniendo en cuenta que en los ensayos de interacción con el DNA de Nf se observó la formación de un complejo con las proteínas p4N y p6N, decidimos analizar también mediante ensayos de protección del DNA a la digestión con DNasa I la conformación de estos complejos. El resultado mostró que en los sitios 1 y 3 de Nf se encuentra unida la proteína p4N y que en la secuencia comprendida entre ambos está unida la proteína p6N (no mostrado). Este es el mismo patrón que producen las proteínas p4 y p6 de  $\phi 29$  en su genoma. Cuando ensayamos en el DNA de Nf lo que ocurría con las proteínas de  $\phi 29$ , vimos que el patrón al que daba lugar el complejo era igual que el producido por las proteínas de Nf, pero se requería 8 veces más cantidad de p6 de  $\phi 29$  para obtenerlo (no mostrado).

### 8.3) Análisis de la competición de la RNAP con los complejos p4-p6, homólogo y heterólogo:

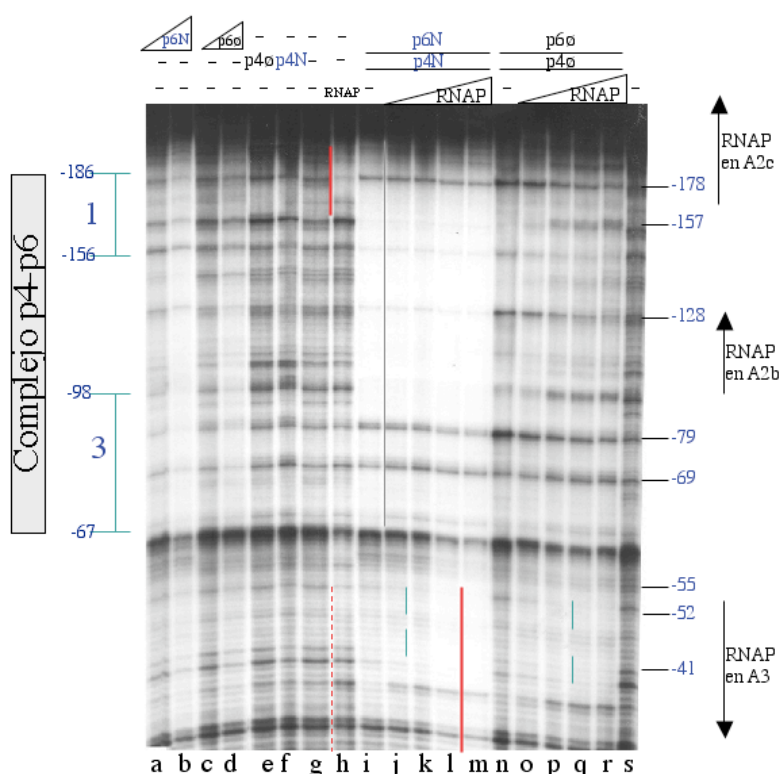
Al igual que en el fago  $\phi 29$ , los genes de Nf son transcritos por la RNAP de *B. subtilis*. Si el sistema de regulación de la transcripción de Nf funcionase de forma similar a como lo hace el de  $\phi 29$ , la unión de la RNAP en los promotores tempranos solaparía con el complejo p4N-p6N descrito en el apartado anterior. Con la intención de obtener más información sobre el proceso de transcripción de Nf, se realizaron ensayos de competición por la unión al DNA de la RNAP y del complejo p4N-p6N (figura 56). El experimento consistió en permitir la formación del complejo p4N-p6N sobre un fragmento de DNA de Nf (+24 del A2c al +14 del A3) marcado radiactivamente y adicionar cantidades crecientes de RNAP, posteriormente someter el DNA a la digestión con DNasa I, y resolver los fragmentos producidos en geles desnaturalizantes de poliacrilamida-urea.

La unión de la RNAP al DNA de Nf en ausencia de otras proteínas se puede observar en el carril h (figura 54), donde aparece unida al promotor temprano A2c (línea roja continua), y protege parcialmente el promotor A3 (línea roja discontinua) (comparar carriles h y g). El complejo p4N-p6N sobre el DNA de Nf (carril i) ocupa la secuencia comprendida desde el sitio 1 al sitio 3 (línea gris). A medida que aumenta la



cantidad de RNAP adicionada (j-m), se observa como el patrón va cambiando en la región del promotor A3, indicando que la unión de la RNAP en este promotor se vuelve más estable en presencia del complejo p4N-p6N. A las dosis más bajas de RNAP aparecen dos zonas protegidas (línea verde, carriles j-k), que podrían deberse a la unión de las subunidades  $\alpha$  de la RNAP. A las dosis más altas se ve la RNAP unida en el promotor A3 protegiendo la secuencia desde la posición -55 hasta el final del fragmento (línea roja). En la región de los promotores tempranos, en cambio, el complejo permanece inalterado ante cualquier dosis de RNAP adicionada, indicando que la RNAP no puede competir con él.

Cuando el complejo previamente formado sobre el DNA de Nf tiene las proteínas p4 y p6 de  $\phi 29$  (línea n) el resultado es diferente. Al ir adicionando cantidades crecientes de RNAP, ésta es capaz de desplazar al complejo p4-p6 formado en la región



de los promotores tempranos (líneas o-q). En la región del promotor A3 se detecta la unión de lo que podrían ser las subunidades  $\alpha$  de la RNAP (en verde, carril r), pero ni a las dosis más altas de proteína se consigue ver la protección producida por la unión de la RNAP al promotor. Este resultado nuevamente indica que el complejo p4-p6 con las proteínas de  $\phi 29$  en el DNA de Nf no puede mantener a la RNAP unida en el promotor A3.

Figura 56: Análisis de la competición de la RNAP y el complejo p4-p6 en el DNA de Nf. El fragmento de DNA de 259 pb se marcó radiactivamente en el extremo 5' de la posición +14 del A3. Las posiciones indicadas en la figura están numeradas desde el promotor A3. Las concentraciones de proteínas que se usaron para formar el complejo p4-p6 fueron p4N y p4 $\phi$  (700 nM), y de p6N y de p6 $\phi$  (20  $\mu$ M); las dosis de RNAP que se utilizaron para competir los complejos fueron 14, 28, 56 y 112 nM (carriles j-m y o-r). Las concentraciones de los controles del ensayo fueron 700 nM de p4N y p4 $\phi$ ; 10 y 15  $\mu$ M de p6N y 15 y 20  $\mu$ M de p6 $\phi$ .

#### 8.4) Regulación de la transcripción *in vitro* del DNA de Nf:

Se analizó la coordinación de la regulación de los promotores tempranos y tardíos de Nf. Para ello se llevaron a cabo ensayos de transcripción “*in vitro*” en presencia y en ausencia de p4N y p6N. Como se puede ver en la figura 57 A, la RNAP, en presencia de un fragmento de DNA de Nf que incluye desde la posición +68 del A2c hasta la +48 del A3 y de los cuatro NTPs (línea a), transcribe los promotores tempranos A2c y A2b. La transcripción desde el promotor A2c da lugar a la aparición de dos transcritos de unos 70 nucleótidos de tamaño, y desde el promotor A2b se transcribe un

RNA de 163 nucleótidos; también hay una pequeña transcripción basal desde el promotor A3 (48 nts de tamaño).

Cuando a este sistema se le adiciona p4N (línea b) se observa la activación del promotor A3; a dosis más altas de p4N se ha visto también represión parcial de los promotores tempranos A2b y A2c. En presencia sólo de la proteína p6N se puede ver que la transcripción desde los promotores tempranos no está prácticamente afectada indicando que no interfiere con la transcripción desde estos promotores (línea d). Cuando se añaden ambas proteínas, p4N y p6N, se observa una represión total de los promotores tempranos y un aumento en la activación de la transcripción tardía (línea c). Esto indica que la regulación simultánea de los promotores A2c, A2b y A3 de Nf requiere la presencia de ambas proteínas, p4N y p6N, al igual que ocurre en el sistema de regulación de  $\phi 29$  con las proteínas p4 $\phi$  y p6 $\phi$ . Para estudiar la posibilidad de que las proteínas de ambos fagos fuesen intercambiables, se analizó la transcripción desde los

promotores de Nf pero con las proteínas de  $\phi 29$  y con las proteínas combinadas de  $\phi 29$  y de Nf. Cuando ambas proteínas eran del fago  $\phi 29$ , no se produjo ni la represión de los promotores tempranos, ni la activación del promotor A3 (líneas g-i). Cuando la proteína p4 era de Nf y la proteína p6 de  $\phi 29$ , el resultado era idéntico a cuando se añadía sólo p4N, es decir, se observa una activación parcial del promotor A3 y no hay efecto sobre los promotores tempranos (comparar líneas b y e); en el caso en el que la proteína p4 era la de  $\phi 29$  y la p6 era de Nf, entonces el comportamiento era igual al de la proteína p6N sola, observándose la ausencia de represión y de activación (carriles d y f).

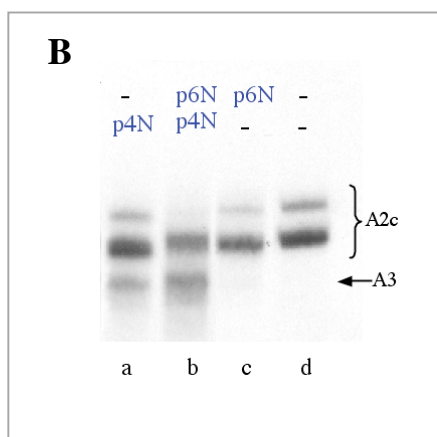
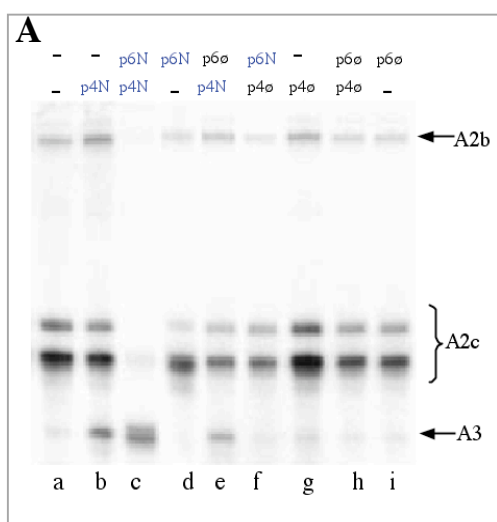


Figura 57: Ensayos de transcripción “in vitro” truncada. En A se presenta el ensayo de la transcripción de los promotores A2c, A2b y A3 de Nf. Las concentraciones de las proteínas utilizadas fueron: 100 nM de p4N y p4 $\phi$ , 10  $\mu$ M de p6N y p6 $\phi$  y 20 nM de RNAP de *B. subtilis*. En B se muestra la transcripción de los promotores A2c y A3 de  $\phi 29$ . Las concentraciones de proteínas usadas fueron 160 nM de p4N, 10  $\mu$ M de p6N y 20 nM de RNAP de *B. subtilis*. Los fragmentos de DNA se incubaron con las proteínas indicadas y las reacciones de transcripción se iniciaron adicionando los cuatro ribonucleótidos, estando el UTP marcado radiactivamente. Los transcritos se resolvieron en geles desnaturizantes al 6% de poliácridamida-8M urea.

Puesto que las proteínas de Nf formaban correctamente los complejos en el DNA de  $\phi 29$ , analizamos la capacidad que tenían para regular la transcripción de los promotores A2c y A3 de  $\phi 29$ . El DNA utilizado fue un fragmento que contenía desde la posición +75 del A2c hasta la +67 del A3 de  $\phi 29$ . Como se puede ver en la figura 57 B, la proteína p4N sola fue capaz de activar el promotor A3 (carril a), el complejo p4N-p6N reprimió ligeramente el promotor A2c y activó algo más el A3 (comparar a, b y c). Por lo tanto, las proteínas de Nf son capaces de regular “in vitro” los promotores de



ø29, aunque no funcionan al mismo nivel que en sus propios promotores. Las proteínas de Nf también fueron capaces de regular el promotor A2b de ø29 (no mostrado).

Estos resultados nos han permitido concluir que ambos fagos comparten un mecanismo de regulación muy similar con algunas diferencias; pero estos mecanismos, a su vez, son exclusivos para cada fago de forma que sus componentes no son intercambiables; las proteínas de ø29 solamente funcionan en el DNA de ø29, mientras que las proteínas de Nf lo hacen en cualquiera de los dos sistemas, aunque a diferente nivel.

## *DISCUSIÓN*

Los resultados obtenidos en este trabajo nos han permitido profundizar en los mecanismos moleculares y estructurales que gobiernan la interacción de una proteína reguladora de la transcripción con sus sitios de unión en el DNA, así como en la formación de un complejo regulador funcionalmente activo y en la conservación evolutiva de dicho sistema de regulación de la transcripción en fagos relacionados.

### **1.- CONFORMACIÓN DEL DNA EN LA SECUENCIA DE LOS PROMOTORES A2c, A2b y A3:**

El análisis de la conformación del DNA de la región 1 de unión de p4 reveló que esta secuencia no tiene curvatura intrínseca. No se detectó una curvatura significativa en el DNA, ni en ensayos de movilidad electroforética a diferentes temperaturas, ni con experimentos de permutación circular en ausencia de proteína. Sin embargo, la proteína p4 es capaz de unirse a los sitios 1 y 2 presentes en esta región y de inducir una curvatura en el DNA similar a la que genera cuando se une a sus sitios de la región 2, donde sí existe una curvatura intrínseca de unos 45° (Rojo *et al.*, 1990).

Estos resultados indican que la proteína p4 no requiere, para su unión, un DNA intrínsecamente curvado. Además, pudimos comprobar con diferentes aproximaciones, que la proteína p4, siempre que se une a la región 1 completa (sitios 1 y 2), genera una curvatura de aproximadamente 86°. Por una parte, se analizó la curvatura inducida por la proteína p4 en fragmentos con mutaciones en la secuencia de la región 1 (Mut 1, Mut 2 y Mut 3); por otra, se analizó la inducción de la curvatura en esta región por proteínas p4 modificadas en residuos implicados en la interacción con el DNA (R54A, K51A, Q5A, H10A y K36A), y por último, se estudió la generación de la curvatura pero con la proteína p4 formando el complejo con p6; en todos estos casos observamos que la p4 induce una curvatura alrededor de 86°.

Mediante procesamiento informático (programa CURVES) de la topología del DNA de la estructura cristalográfica del complejo p4-DNA, determinamos que la unión de p4 induce una curvatura de 54,2° en la secuencia de DNA correspondiente a los 31 pb del sitio 3 y que está aproximadamente centrada en la mitad del sitio de unión. Si asumimos que esto ocurre de forma similar en el resto de los sitios de unión, la curvatura inducida de 86° es el resultado de la unión simultánea de p4 a los dos sitios de cada región. Esperaríamos entonces, que el centro aparente de dicha curvatura se localizase entre ambos sitios de unión; en cambio, cuando se analizó el centro de la curvatura inducida por p4 en la región 1 observamos que estaba localizado en el sitio 1 un poco desplazado hacia el sitio 2, en concreto, a 10 nucleótidos de distancia del centro entre ambos sitios. El centro de curvatura en el DNA de la región 2 tampoco está centrado entre los dos sitios de unión, sino que se encuentra en el sitio 4 en su parte más próxima al sitio 3 y a una distancia, también, de 10 bases del centro entre ambos (Rojo *et al.*, 1990). La razón del desplazamiento del centro de curvatura con respecto al esperado podría deberse a que, como se puede ver en la estructura del complejo p4-DNA, cuando la p4 se une al sitio 3, genera una curvatura no planar, apareciendo una ligera torsión del DNA en el eje Z (figura 58). La presencia de esta desviación, al tratarse de un cambio conformacional en el DNA, se refleja en la movilidad electroforética del complejo a bajas temperaturas en geles de poliacrilamida nativos.

El hecho de que en ambas regiones el centro de curvatura se encuentre a la misma distancia del centro entre los dos sitios contiguos, implica que probablemente el centro de la curvatura final de 86°, que se induce en el DNA de cada región, esté centrado entre ambos sitios.

Cuando se analizó mediante ensayos de permutación circular el DNA de la región 1 con el sitio 1 mutado (Mut 1) observamos que el centro de curvatura se desplazaba del sitio 1 al sitio 2 y que la unión de p4 a este sitio 1 modificado era débil e inestable. Esta inestabilidad en la unión de p4 al sitio 1 podría estar distorsionando la

movilidad electroforética del complejo y, en consecuencia, cambiar el centro de la curvatura del DNA (17 pb con respecto al salvaje y solo 8 con respecto al centro entre los sitios).

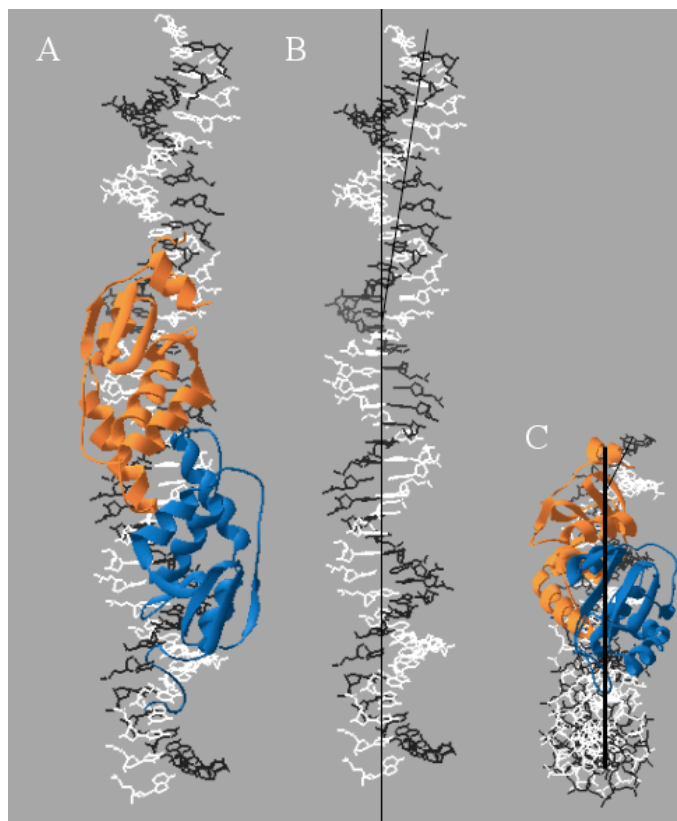


Figura 58: En A se muestra un dibujo del complejo p4-DNA, obtenido por cristalografía, visto desde arriba. En B se ha eliminado la proteína para poder apreciar la topología del DNA; se dibujó sobre la cadena una línea negra a lo largo del eje del DNA y se puede ver como se desvía el eje de la cadena. En C se muestra un dibujo de la misma estructura desde una visión frontal inclinada hacia arriba; de este modo se puede ver el cambio en la dirección que experimenta el eje del DNA tras la unión de la proteína.

## 2.- INTERACCIÓN DE p4 CON SUS SITIOS DE UNIÓN

### - Sitio mínimo de unión de p4:

En estudios previos, llevados a cabo en nuestro laboratorio, sobre la unión de la proteína p4 al DNA se había descrito que un sitio de unión incluía la secuencia con la repetición invertida “ 5´AAAA(T/A)GTT 3´” (Barthelemy y Salas, 1989); también se definió la importancia de las guaninas presentes, cuya metilación interfería con la unión de p4 (Barthelemy y Salas, 1989; Monsalve *et al.*, 1998) y que, en ensayos de protección del DNA a la ruptura por radicales hidroxilo en un DNA con la secuencia de las regiones 1 y 2, p4 protegía grupos de nucleótidos llamados huellas (Camacho y Salas, 2004). Eliminando sucesivamente 1, 2, 3 ó 4 de estas huellas en fragmentos de DNA que contenían una región completa de unión de p4, pudimos comprobar que cada tres de estas huellas sucesivas, y coincidentes con una repetición invertida, eran la secuencia mínima requerida por p4 para su unión al DNA. De estos ensayos pudimos concluir que la proteína p4 se unía de forma estable e individual a los sitios 1, 2 y 3, mientras que no lo hacía en el sitio 4. También pudimos establecer la afinidad que la p4 tenía por sus sitios de unión, presentando el doble de afinidad por el sitio 3 que por el sitio 1 y cinco veces más afinidad por el sitio 1 que por el sitio 2.

- Unión de p4 al DNA:

La resolución de la estructura cristalográfica de p4 unida al sitio 3 permitió determinar que p4 se une a su sitio en el DNA en forma de dímero. La proteína tiene en su extremo amino terminal una estructura única con forma de anzuelo a la que se denominó Garfio. Esta estructura sobresale hacia el exterior en ambos extremos del dímero de p4 y se introduce en el surco mayor del sitio de unión; es el elemento más importante en el reconocimiento e interacción de p4 con el DNA.

En la estructura cristalográfica se observó que 9 residuos de cada monómero podrían estar estableciendo interacciones con el DNA, y de éstos solamente dos lo harían de forma específica, entendiendo como tal aquellas interacciones que tienen lugar entre la cadena lateral de un aminoácido y la base nitrogenada de un nucleótido. Estos dos residuos son la Arg6 y la Gln5.

Arginina 6: La cadena lateral de este aminoácido establece, por monómero, dos interacciones fuertes, a una distancia menor de 3Å, con las guaninas en posición 13 del sitio de unión. Este residuo está localizado en el Garfio y está orientado hacia el exterior de la molécula. Las dos guaninas están conservadas en todos los sitios de unión de p4 excepto en el de menor afinidad, el sitio 4, que solo posee la correspondiente a la posición -13.

En el desarrollo de este trabajo hemos podido definir que la interacción Arg6-Guanina(13) es el determinante principal del *reconocimiento específico directo* de la unión de p4 a sus sitios. En primer lugar, la adición de una guanina al sitio 4, en la correspondiente posición +13, permite la formación de un complejo estable de la proteína p4 unida a un fragmento de DNA con la secuencia únicamente del sitio 4. Por otra parte, la unión de p4 al sitio 3, que es por el que presenta mayor afinidad, se ve reducida entre 10 y 30 veces cuando modificamos por adeninas las guaninas en posición -13 y +13, respectivamente. También se llevó a cabo la caracterización funcional de la proteína mutante R6A y observamos que tiene muy reducida su capacidad de interacción con el DNA, así como la de formar complejos con las otras proteínas del sistema regulador, y por último, vimos que esta proteína mutada era incapaz de regular la transcripción de los promotores A2c, A2b y A3 de  $\phi$ 29.

Glutamina 5: Este residuo también está localizado en el Garfio de p4. En la estructura del complejo p4-DNA se puede ver que la distancia entre la cadena lateral de este aminoácido con la base nitrogenada de la timina en posición +15 es de 3,04 Å, suficiente para una interacción fuerte entre proteína y DNA, mientras que con la timina en posición -15 la distancia es de 3,32Å. Quisimos comprobar la importancia que, presumiblemente, tenía esta interacción Gln5-Timina(15) y lo abordamos de distintas formas. En primer lugar, puesto que sabíamos que p4 tiene mayor afinidad por el sitio 3 que por el sitio 1, y viendo que éste carecía de la timina en posición -15, modificamos el sitio 1, con la intención de aumentar la afinidad de p4 por este sitio, insertando una timina en posición -15 (Sit 1A); el resultado mostró que este cambio no aumentaba la unión de la proteína p4 a este sitio 1 modificado. Por otro lado, se generó el mutante Q5A de p4 con el objetivo de estudiar la relevancia de este residuo en la interacción de la proteína con su sitio de unión; la proteína p4 modificada se unía bien al DNA, formaba los complejos con las otras proteínas del sistema regulador igual que la p4 salvaje y regulaba la transcripción al mismo nivel que lo hacía la proteína p4 no modificada. Todos estos datos, junto con el hecho de que la proteína p4 del fago Nf, con la que comparte un 78,5% de identidad, tenga en posición 5 una alanina en lugar de una glutamina sugería que la interacción Gln5-Timina(15) deducida de los datos cristalográficos no era funcional.

Para descartar definitivamente la funcionalidad de la Gln5 de p4, se llevó a cabo, en colaboración con el Dr. Jesús Mendieta, el análisis de la distancia entre los átomos implicados en la interacción Gln5-T15, en una trayectoria de 5 nanosegundos de una Simulación Dinámica Molecular del complejo p4-DNA. Como se puede ver en la figura 59, en ambos monómeros, la distancia entre el Ne de la glutamina y el O<sub>4</sub> de la timina es de unos 6 Å a lo largo de la trayectoria; esto implica demasiada distancia para que se pueda formar un enlace de hidrógeno entre ellos, sugiriendo, por tanto, que esta interacción no es necesaria para la estabilización del complejo p4-DNA; puesto que el complejo está formado a lo largo de la simulación y según estos datos, la interacción Gln5-T(15) no se estaría produciendo. Este resultado está de acuerdo con el resto de ensayos realizados y nos hace pensar que la disposición y proximidad de la glutamina y la timina en la estructura p4-DNA es una consecuencia del proceso de cristalización.

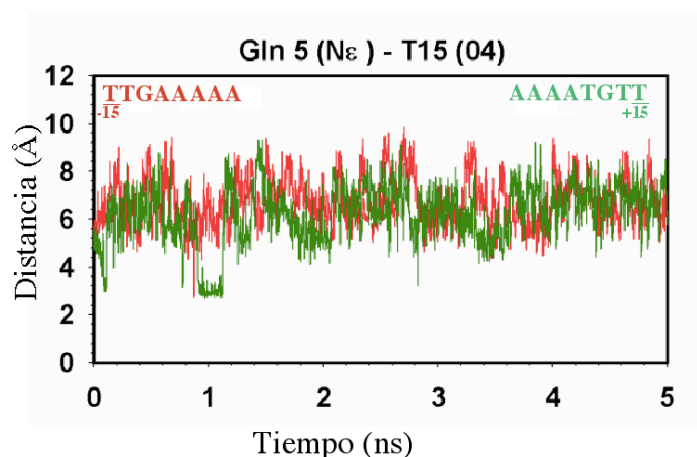


Figura 59: Perfiles dinámicos de las distancias entre el Ne de la glutamina en posición 5 de cada monómero de p4 y el O<sub>4</sub> de las timinas en posición 15 de los extremos del sitio 3, frente al tiempo (5 nanosegundos). Ensayo de simulación dinámica molecular llevado a cabo por el Dr. Jesús Mendieta (Mendieta, Pérez-Lago, Salas y Camacho, enviado a publicar).

Si las interacciones de las Gln5 con las timinas en posición 15, realmente no se producen, y teniendo en cuenta que ningún otro aminoácido establece interacciones con los nucleótidos en posiciones 14 y 15, podemos concluir que un sitio de unión de p4 tiene 27 pb y no 31 pb como veníamos suponiendo. La secuencia consenso sería: **CTTTT-X<sub>8</sub>-TT-X<sub>7</sub>-AAATG**

El reconocimiento *específico directo* del sitio de unión de p4, por lo tanto, se basa exclusivamente en las interacciones Arg6-G(13); de este modo, la p4 se podría unir



Figura 60: Foto de pantalla con el resultado obtenido tras el análisis de la secuencia del genoma de ø29 con el programa “Phi-29 DNA Strand Counter”.

a cualquier secuencia que tuviese dos guaninas localizadas en cadenas complementarias y separadas por 25 pares de bases. Diseñamos, en colaboración con el analista de sistemas D. Ángel Sánchez Enano, un programa informático que nos permitió calcular el número de veces que se repetía este evento en el genoma de ø29; el resultado mostró que en 781 ocasiones una citosina se encuentra a una distancia de 25 bases de una guanina en la misma cadena de DNA (figura 60). Por lo tanto, el resto de interacciones que establecen los aminoácidos de p4 con el esqueleto fosfato del DNA deben estar aportando, a su vez, especificidad al reconocimiento y unión de p4 a sus sitios.

Cuando se estudió, mediante mutagénesis puntual dirigida de la proteína p4, la importancia que tenían en la unión al DNA los residuos implicados en la interacción con el esqueleto fosfato, observamos que la modificación de la Tyr33, de la Thr4 o de la Lys76, daba lugar a proteínas cuya capacidad de unión al sitio 3 era prácticamente nula. En la estructura cristalográfica del complejo estos residuos establecen interacciones con el esqueleto fosfato del DNA a ambos lados de las secuencias de pares de bases A/T localizadas entre las posiciones 8 y 12 del sitio 3 de unión (solo uno de los monómeros interacciona mediante Lys76) (figura 61). La His10 de p4 interacciona con el esqueleto fosfato en la misma posición que la Thr4, pero su modificación a alanina no produce un efecto tan drástico en la unión al DNA.

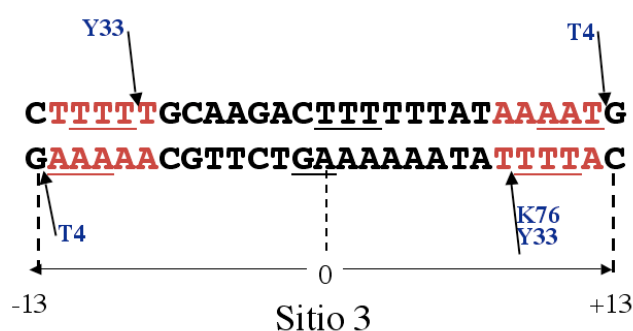


Figura 61: Secuencia del sitio 3. Se presenta en rojo la secuencia de pares de bases A/T del sitio 3 donde los residuos Thr4, Lys76 y Tyr33 establecen las interacciones con el esqueleto fosfato del DNA. Con líneas horizontales se señalan las huellas que se obtienen en ensayos de ruptura química por radicales hidroxilo.

Por otra parte, debemos considerar que la treonina 4, además de interactuar con el DNA, establece enlaces con otros aminoácidos que forman parte del Garfio (figura 21), de modo que no se puede descartar que el fuerte defecto en la interacción con el DNA que presenta la proteína mutante T4A, se deba a una desorganización de este motivo estructural ya que es, precisamente en el Garfio, donde se encuentra el residuo Arg6 que hemos descrito como imprescindible para el reconocimiento del sitio de unión.

Las secuencias de pares de bases A/T de los extremos del sitio presentan varias características que podrían incidir en la interacción de p4 con el DNA. En primer lugar, son dos de los tres grupos de nucleótidos que resultan protegidos por la p4 en ensayos de ruptura química del DNA con radicales hidroxilo y que hemos denominado *huellas*. Por otra parte, estas secuencias tienen el surco menor estrechado en la estructura cristalográfica del complejo p4-DNA (en verde, figura 62).

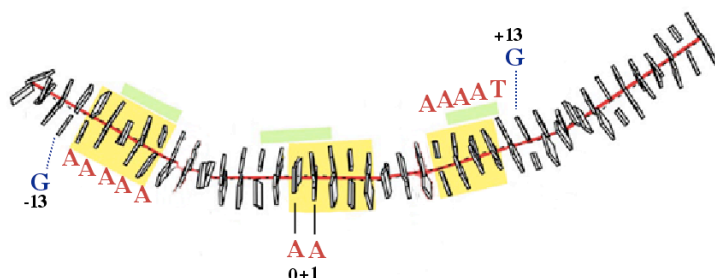


Figura 62: Esquema de la conformación del DNA de 41 pb con el que se cristalizó la proteína p4. En amarillo se muestran los grupos de adeninas consecutivas que aparecen en la secuencia del sitio 3, y en verde las secuencias con el surco menor estrechado en el cristal del complejo p4-DNA. (Badía *et al.*, 2006).

Y, por último, el cambio en el sitio 3 de las adeninas en las posiciones 10 y 11 por citosinas, produce la pérdida prácticamente total de la interacción de p4 con el DNA. La proteína p4, en cambio, recupera su capacidad de interacción con el DNA cuando en lugar de modificar las adeninas a citosinas, se modificaban a timinas. Estos resultados nos hicieron pensar que la ruptura de este grupo de pares de bases A/T era el

responsable de la pérdida de la interacción. Llevamos a cabo, entonces, modificaciones con análogos de nucleótidos en la secuencia del sitio 3 y pudimos demostrar que la importancia de estos grupos de adeninas y timinas radica en el hecho de ser secuencias deformables; la modificación de adenina a DAP que mantiene la secuencia de A/T no restablece la unión de p4, mientras que la modificación de adenina a inosina, que rompe la secuencia de A/T por su apareamiento con la citosina, sí permitía la unión de p4. La diferencia más importante entre estos análogos de nucleótidos es que el par I:C es más deformable que el par DAP:T.

Las mutaciones de residuos de la proteína con el efecto más drástico interaccionan con el esqueleto fosfato de esta región de pares de bases A/T y probablemente están contribuyendo, a través de sus interacciones, a deformar el DNA amoldándolo para que pueda tener lugar la unión de forma estable. La falta de alguno de estos residuos implica la pérdida de interacciones importantes para acomodar a la proteína en su sitio de unión. Asimismo, la disminución de la flexibilidad de la secuencia y la pérdida de su capacidad para deformarse hacen que proteína y DNA no puedan acoplarse adecuadamente y por lo tanto, la unión se vea impedida. De estos datos podemos concluir que, en gran medida, el mecanismo de interacción de p4 con su sitio de unión está mediado por un proceso de reconocimiento *específico indirecto*, puesto que se produce un efecto muy drástico tras la modificación de dos pares de bases en la secuencia (en posiciones 10 y 11) con las que aparentemente no interacciona la proteína.

En la secuencia del sitio 3, como se puede ver en las figuras 61 y 62, existe una tercera región protegida por la proteína a la ruptura con radicales hidroxilo y que también coincide con la tercera zona donde el surco menor está estrechado en la estructura. Las posiciones 0 y +1 de esta secuencia son pares de bases A/T en los dos sitios por los que p4 tiene mayor afinidad, el sitio 3 y el sitio 1; en cambio, en el sitio 2 son pares de bases G/C; cuando se cambiaron ambas posiciones a A/T (Sit 2A), aumentó unas 8 veces la afinidad de p4 por este sitio 2 modificado. Este resultado de nuevo confirma la importancia de la flexibilidad que debe poseer el DNA de un sitio de unión de p4 en determinados puntos de la secuencia para adaptar su conformación a la interacción de la proteína. Esta secuencia coincide con el lugar donde los residuos Arg54 y Lys51 establecen sus interacciones con el esqueleto fosfato del DNA. La modificación de estos aminoácidos a alanina da lugar a proteínas funcionalmente activas y, aunque en algunas ocasiones se requieren en mayor concentración, en general, su comportamiento es muy similar a la p4 silvestre. Los aminoácidos Lys51 y Arg54 están localizados en la región de dimerización de la proteína y establecen interacciones que, de forma individual, no son muy importantes, en el complejo p4-DNA. Probablemente, su papel sea el de estabilizar el DNA fijando la zona central del sitio al dímero, de modo que las interacciones que establece uno de estos residuos serían suficientes para mantener unido el DNA por la zona central.

La pérdida de las interacciones de uno de los residuos que en condiciones óptimas parece inocua, se vuelve imprescindible cuando se modifica la longitud del sitio de unión. Cuando analizamos la influencia que tenía en la unión de p4 la longitud del sitio de unión, vimos que afectaba más la disminución que el aumento de una base en el sitio 3, pero en ambos casos se formaba el complejo con la proteína p4. Cuando estos mismos fragmentos se ensayaron con las proteínas modificadas R54A y K51A, el resultado fue mucho más contundente, puesto que se reducía drásticamente la formación del complejo cuando el DNA tenía una base menos que el sitio 3 silvestre. Puesto que, para que pueda tener lugar la interacción con el DNA del sitio es necesario que los Garfios se introduzcan por los surcos mayores correspondientes y se establezcan las



interacciones de las argininas 6 con las guaninas en posición 13, pensamos que en un DNA con una base menos, probablemente se produce una tensión que dificulta mucho la estabilidad de la proteína. En el caso de la p4 salvaje disminuye entre 3 y 10 veces la formación de complejos. En estas condiciones las interacciones de la región central (Lys51 y Arg54), no se pueden establecer de forma correcta y el hecho de que un aminoácido esté modificado perdiendo sus interacciones, hace que el otro no sea suficiente para acomodar el DNA en el complejo correctamente. Cuando el DNA posee un par de bases más, también se observa un efecto mayor con la proteína careciendo de alguno de los dos residuos que con la proteína salvaje, pero es menos drástico.

Por tanto, aparte de las guaninas en posición 13 con las que interaccionan las argininas 6 de los Garfios de la proteína, en la secuencia de un sitio de unión de p4 existen tres zonas muy importantes para que la interacción p4-DNA se produzca de forma adecuada: la secuencia de la región central y las de los dos extremos del sitio de unión. Estas tres zonas comparten como características comunes que están formadas por pares de bases A/T, que son deformables y que varios residuos de la proteína establecen interacciones con el esqueleto fosfato de las mismas.

Cuando se analiza la secuencia de los extremos del sitio de unión se puede ver que no es una repetición invertida perfecta, y al estudiar la influencia de las diferencias observamos que la afinidad de p4 por un sitio 3 modificado que posee sus dos extremos iguales, siendo estos 5'AAAATG 3', era cuatro veces mayor que la afinidad por un sitio de unión con sus dos extremos con la secuencia 5'AAAAAG 3'. De estos ensayos se dedujo que p4 no tiene la misma afinidad por los dos extremos y que los dos

monómeros de p4 no se unen igual al DNA. Denominamos monómero A al que se une más débilmente y monómero B al que lo hace con más fuerza. Para confirmar estos resultados, analizamos la formación de complejos de p4 con sitios 3 que tenían modificadas cada una de las guaninas en posición 13 por adeninas. El resultado mostró que la mutación en el extremo al que se une el monómero B (G+13), que es el más fuerte, afectaba más a la interacción de la proteína que la modificación de la guanina -13; se confirma, de este modo, que la interacción de un monómero es más fuerte que la del otro. La razón por la cual una timina, en lugar de una adenina, en posición +12, localizada después de una secuencia de adeninas consecutivas en el sitio de unión, favorece la interacción de p4, podría estar en el hecho de que un paso pirimidina-purina (T/G) es más susceptible de deformación

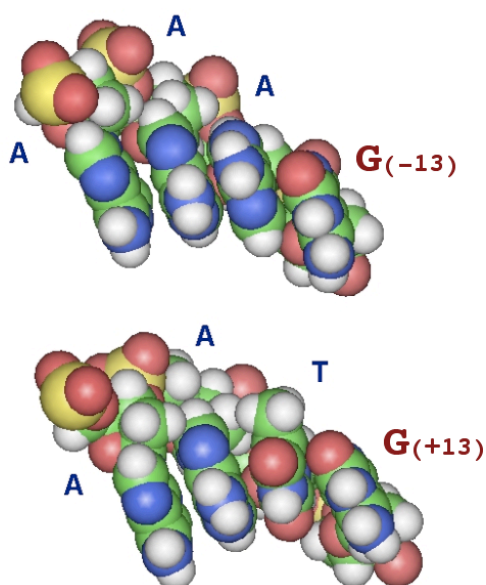


Figura 63: Posición y estructura de las bases que forman las repeticiones invertidas en los dos extremos del sitio 3 de unión de p4. Se puede ver mayor superficie de contacto entre dos purinas consecutivas que entre una pirimidina y una purina.

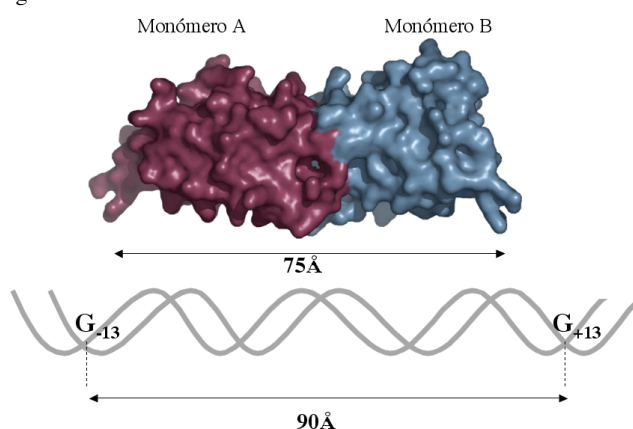
que un paso purina/purina (A/G), porque en el primer caso hay una menor superficie de las bases solapando (figura 63) (Dickerson y Chiu, 1997; Olson *et al.*, 1998), de modo que podría estar favoreciéndose la orientación de la guanina en posición 13 para la interacción con la arginina 6 del monómero B.

Además, en la estructura del complejo de la proteína unida al DNA, se puede observar una posible interacción, mediada por agua, entre la arginina 77 del monómero B con la timina en posición +12, ya que se encuentran a una distancia de 2,78Å, que podría estar aportando estabilidad a la interacción del monómero B con este extremo del sitio 3 de unión.

El hecho de que los dos monómeros de p4 no se unen con la misma afinidad al DNA se pudo también confirmar también en otros sitios de unión de p4, llevando a cabo la mutagénesis de cada una de las guaninas de los sitios de unión, en fragmentos de DNA conteniendo dos de estos sitios. Cuando se llevaron a cabo ensayos de interacción con estos fragmentos, pudimos ver que para cada sitio, la pérdida de una guanina o de la otra no tenía el mismo efecto. En el sitio 1, cuando la guanina modificada era la que estaba en posición -13 (G11), se disminuía la formación de complejos entre 4 y 12 veces, mientras que, en el caso de la guanina en posición +13 la disminución era entre 3 y 5 veces (G12). En el sitio 2, la modificación de la guanina en posición +13 (G22) no tenía ningún efecto en la formación de complejos con el DNA de la región 1, mientras que la modificación de la guanina en posición -13 (G12) producía una disminución en la formación de complejos de entre 2 y 3 veces. En ambos sitios, el extremo que produce mayor efecto posee un paso purina/pirimidina, lo cual apoyaría la hipótesis de que el paso G/T esté favoreciendo el establecimiento de las interacciones p4-DNA. En el caso del sitio 3 pudimos corroborar que el extremo donde está la guanina en posición +13 era el más fuerte, puesto que la mutación G32 implicaba una pérdida de al menos 30 veces en la capacidad de formar complejos, mientras que la modificación en -13 (G31) apenas afectaba a la interacción con el DNA de la región 2. La inserción en el sitio 4 de una guanina en la correspondiente posición +13 mejora la unión de p4 a un DNA que posee únicamente la secuencia de este sitio; cuando se analizó la unión de p4 a toda la región 2 (sitios 3+4) ninguna de las modificaciones realizadas en el sitio 4 (G41 y G4+) afectaba al complejo p4-DNA, probablemente debido a que la presencia del sitio 3 está enmascarando el efecto de las modificaciones en el sitio 4.

El hecho de que la interacción de cada uno de los monómeros de p4 con el DNA tenga diferente afinidad, está sugiriendo que el mecanismo de unión de p4 implica la interacción inicial del monómero B con la secuencia 5'AAATG 3' y posteriormente la interacción del monómero A con la secuencia 5'AAAAG 3'. Según esto, sugerimos un modelo del mecanismo de unión, que hemos denominado “de cremallera” y que se describe a continuación:

Figura 64A



Se representa el DNA en gris y el dímero de p4 en granate y azul. La distancia entre las Arg6 del dímero de p4 es de 75Å mientras que la distancia entre las guaninas 13 de los surcos mayores es de 90Å; sabemos que la conformación del dímero de p4 en solución apenas presenta cambios con respecto a la estructura de p4 en el complejo con el DNA, de modo que, a priori ambos Garfios no van a

poder introducirse a la vez en los surcos mayores correspondientes, puesto que están muy alejados (figura 64A). Sabemos también que en el cristal, donde el DNA está curvado, ambos Garfios interaccionan simultáneamente con el DNA.

Figura 64B

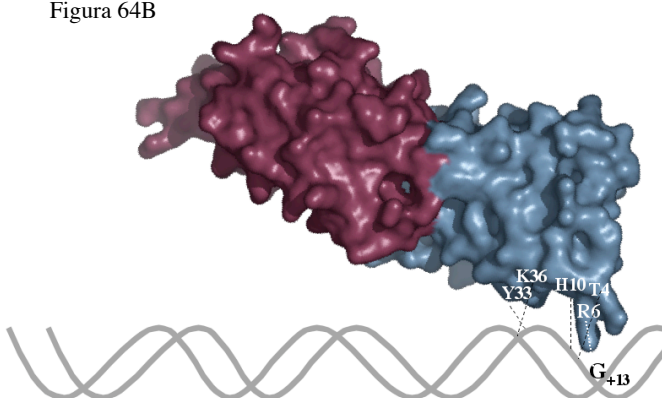


Figura 64C

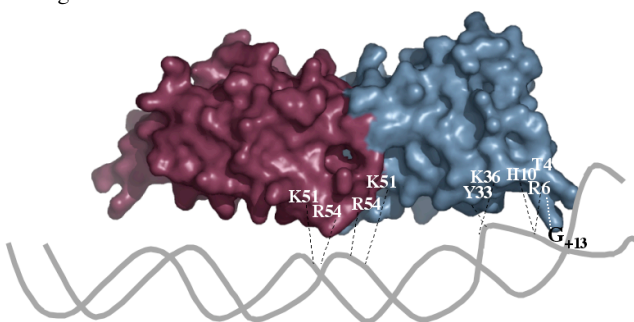


Figura 64D

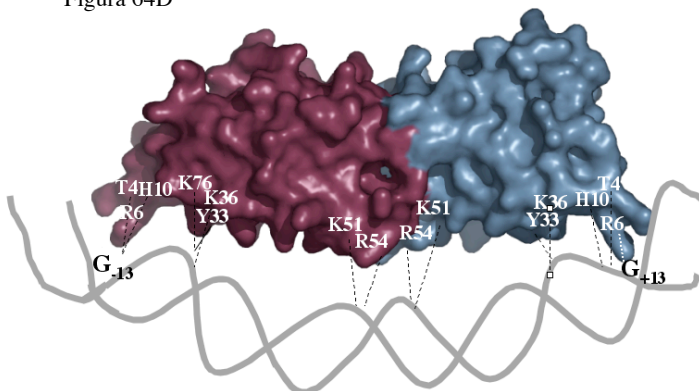


Figura 64: Dibujo que esquematiza el modelo propuesto para el mecanismo de interacción de un dímero de p4 en un sitio de unión del DNA de  $\phi 29$ .

Teniendo en cuenta que entre las proteínas p4 de Nf y de ø29 existe una identidad de un 78,5% y que todos los aminoácidos de la p4 de ø29 implicados en la interacción con el DNA están conservados en la p4 de Nf, podemos asumir que la p4 de Nf también se une como dímero a sus sitios y que probablemente interaccione con el DNA a través de los mismos residuos. Si analizamos la secuencia de los sitios de unión de la p4 de Nf y observamos las características que hemos descrito como determinantes para la unión de la p4 de ø29 a sus sitios (figura 65), podemos observar que prácticamente cumple todas las condiciones de un sitio de unión de la p4 de ø29. La diferencia más importante está en el sitio 3, que posee un nucleótido más, siendo la separación entre las guaninas de 26 nucleótidos. Esta mayor longitud es, probablemente, la responsable de que la afinidad de unión de la p4 de ø29 por el sitio 1 de Nf sea mayor que por el sitio 3 de Nf; además, puesto que la p4 de Nf posee también menor afinidad por el sitio 3 que por el sitio 1 de su genoma, y teniendo en cuenta la gran homología entre ambas proteínas, posiblemente la causa de esta diferencia en la afinidad sea también el par de bases adicional.

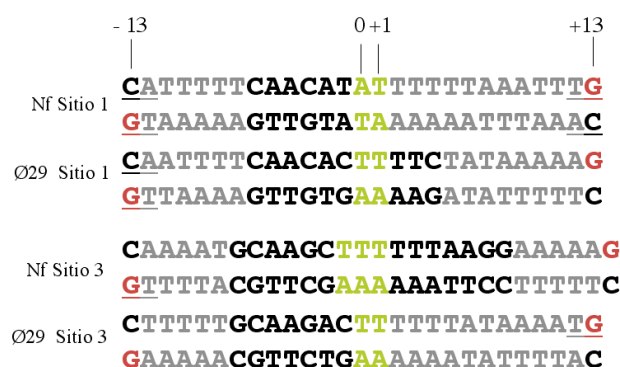


Figura 65: Análisis comparativo de las secuencias de los sitios 1 y 3 de ø29 y de Nf. Se marcan en rojo las guaninas que en ø29 interaccionan con las argininas 6 y sus homólogos en Nf; se presentan en gris las secuencias de pares de bases A/T consecutivas que aportan flexibilidad al sitio de unión y en verde las posiciones centrales de pares de bases A/T con las que interaccionan en ø29 los residuos de la región de dimerización, R54 y K51, y sus posibles homólogos en Nf.

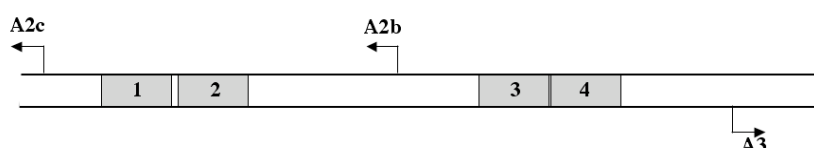
Los dos sitios de unión de la p4 de Nf poseen las guaninas en la correspondiente posición 13. A continuación de las guaninas hay secuencias de pares de bases A/T que podrían tener, al igual que en ø29, capacidad para deformarse, lo que conferiría gran flexibilidad al sitio de unión; los nucleótidos de contacto con los residuos de los extremos del dímero, descritos para la p4 de ø29, implicados en la interacción, se mantienen conservados como pares de bases A/T en los sitios de unión de Nf. También los nucleótidos de la región central del sitio, que interaccionan con los aminoácidos de la zona de dimerización, son pares de bases A/T como en ø29 (figura 65). De todo esto podríamos deducir que la p4 de Nf puede compartir el mismo mecanismo de unión que el que hemos propuesto para ø29; interaccionando primero con la secuencia, a través del Garfio de un supuesto monómero B, y posteriormente aproximando y acomodando el DNA a la proteína, mediante interacciones con el esqueleto fosfato y terminar con la introducción del segundo Garfio en el surco mayor del otro extremo. En el caso del sitio 3 de Nf, por comparación de secuencia y extrapolación de los datos obtenidos para ø29, el monómero B sería el que interaccionaría con el extremo donde está localizada la guanina en la correspondiente posición -13, ya que posee un paso purina/pirimidina; en cambio, curiosamente, ambos extremos del sitio 1 de Nf, que es precisamente por el que la p4 presenta mayor afinidad, tienen el paso GT, no pudiendo establecer a priori con cual de ellos interaccionaría primero la p4N.

### 3.- FUNCIÓN DE LOS SITIOS DE UNIÓN p4 EN ELDNA:

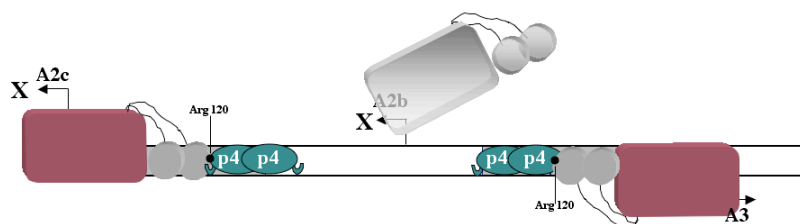
La proteína p4 presenta diferente afinidad por cada uno de sus cuatro sitios de unión. En cada región, donde los sitios se agrupan por parejas, hay un sitio fuerte y otro débil. Utilizamos como herramienta la mutagénesis de las guaninas en posición 13 de cada uno de los sitios, para debilitar la unión del dímero de p4 a sus sitios de unión y así estudiar la implicación de cada uno de ellos en la regulación del paso de la transcripción temprana a la tardía en el fago ø29.

Analizamos los cambios producidos en el sistema regulador de la transcripción tras cada modificación en el DNA, tomando como referencia los complejos formados y los transcritos producidos *in vitro* en el sistema con el DNA salvaje y que se representa en la figura 66.

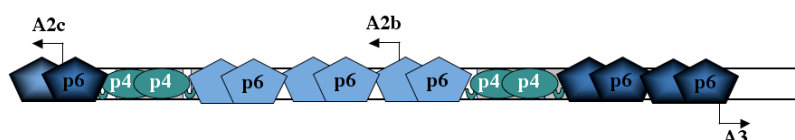
#### A) Organización de los sitios de unión de p4



#### B) Esquema del complejo *in vitro* p4-RNAP.



#### C) Esquema del complejo *in vitro* p4-p6



#### D) Esquema del complejo *in vitro* p4-p6-RNAP

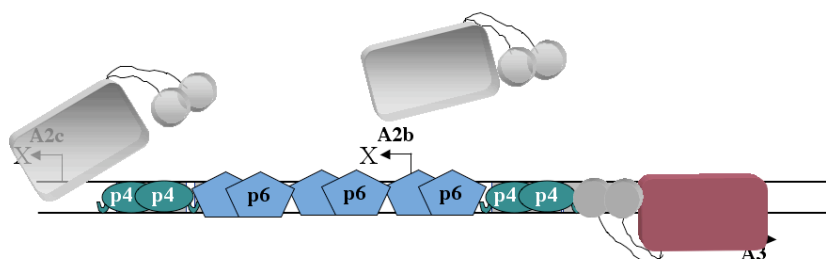


Figura 66: Representación de la formación de los complejos que genera *in vitro* la proteína p4 con la RNAP y/o con la proteína p6, en la secuencia de los promotores A2c, A2b y A3. En A se representa el DNA con los tres promotores y los cuatro sitios de unión. En B, el complejo p4-RNAP. En C, el complejo p4-p6 y en D, el complejo p4-p6-RNAP. La RNAP está representada en granate y gris, la proteína p4 en verde y la p6 en azul. En azul oscuro, la localización de la p6 cuando se adiciona a altas concentraciones.



Tras estudiar la formación de complejos y la transcripción truncada *in vitro* de los DNAs con las guaninas del sitio 1 modificadas, observamos que estas modificaciones afectan mucho a la interacción de la p4 con la secuencia de la región 1. La formación de complejos de proteína p4 con p6 no se produce de forma adecuada, lo cual se traduce en una falta de regulación de la transcripción mediada por dicho complejo; de modo que no se reprime la transcripción temprana ni se activa la tardía más allá de lo que lo hace la proteína p4 sola. La figura 67 representa de forma esquematizada lo que podría estar sucediendo. La falta de interacción de la p4 con el sitio 1 no permite la formación adecuada del complejo con la proteína p6.

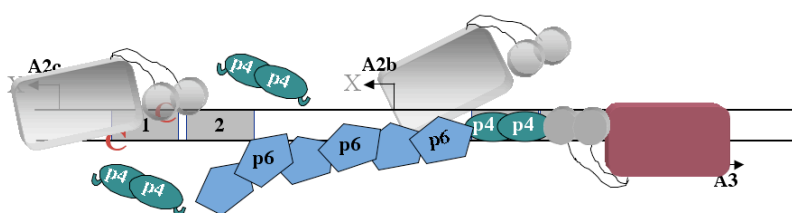


Figura 67: Dibujo que representa los resultados obtenidos tras modificar las guaninas del sitio 1. C indica el lugar de la mutación.

En el caso de los cambios en el sitio 2, el DNA con la guanina en posición +13 modificada (mutante G22) tuvo un comportamiento igual al del DNA no mutado. En cambio, en el mutante G21, la deficiente unión de p4 al sitio 2 afectaba a la capacidad para formar el complejo p4-RNAP y esto influía en la represión mediada por p4 sola del promotor temprano A2c. Podríamos esquematizar el efecto de estos cambios como se muestra en la figura 68.

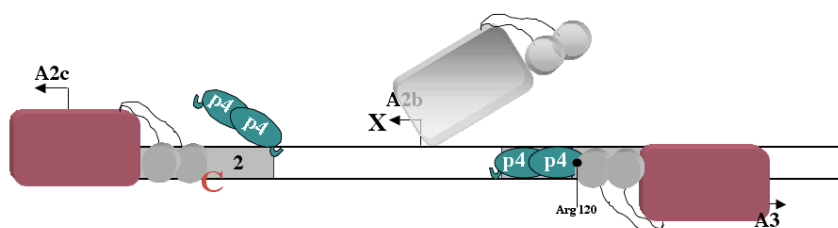


Figura 68: Esquema que representa el efecto obtenido tras las modificaciones en el sitio 2. C indica el sitio donde está la mutación.

Las modificaciones de las guaninas del sitio 3 impidieron la correcta formación del complejo p4-p6 y la consiguiente pérdida de regulación de la transcripción. Pudimos observar que, aunque en el caso de la mutación G31 no había un efecto significativo en la interacción de la proteína con la región 2, la unión de p4 al sitio 3 en este DNA modificado no debe ser apropiada, puesto que ni se reprime adecuadamente la transcripción temprana ni se activa la tardía. En el esquema de la figura 69 se resume lo que podría estar sucediendo. La falta de unión de p4 en el sitio 3 impide la estabilización de la RNAP en el promotor A3 al mismo tiempo que permite su unión en el promotor A2b; además, el complejo p4-p6 no puede formarse adecuadamente.

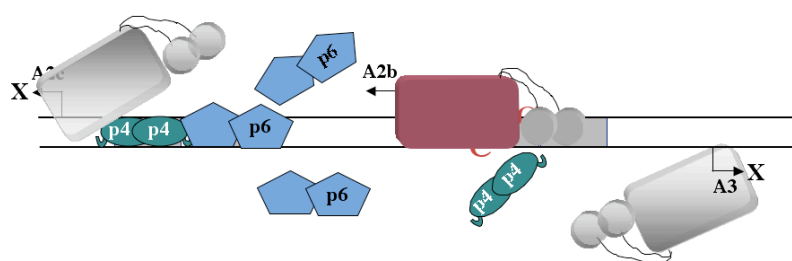


Figura 69: Esquema que representa el efecto que producen las guaninas modificadas del sitio 3. C indica el lugar de la mutación.

Teniendo en cuenta estos resultados y el hecho de que el DNA modificado en el sitio 4 tuviera el mismo comportamiento que el DNA silvestre y asumiendo que *in vivo* la entidad funcional que regula la transcripción es el complejo p4-p6, podemos deducir que los sitios 2 y 4 no tienen un papel directo en la regulación del paso de la transcripción temprana a la transcripción tardía. Además, hemos visto que el mecanismo de regulación de la transcripción de Nf es básicamente igual al de  $\phi 29$ , y este fago carece en su genoma de secuencias homólogas a los sitios 2 y 4 de  $\phi 29$ .

El conjunto de todos estos datos nos llevó a pensar que los sitios 2 y 4 podrían ser fruto de duplicaciones generadas en el DNA, que se han mantenido en el genoma debido a que su presencia sea de alguna forma beneficiosa para la supervivencia del fago  $\phi 29$ . Quizás estos sitios de unión estén desempeñando un papel de almacén o de reservorio de la proteína p4. La función de p4 es crucial en el ciclo vital del fago, puesto que es la responsable de regular la expresión de sus genes. La transcripción temprana es la encargada de la producción de genes implicados, básicamente, en la replicación y generación de nuevos genomas completos; para que éstos puedan ser infectivos deben introducirse en las cápsidas y formar finalmente las partículas virales, lo cual depende de la transcripción tardía. Esto exige alcanzar un equilibrio entre la represión y la activación de estas dos etapas transcripcionales, y es la regulación mediada por p4 la principal responsable de conseguirlo. La presencia de los sitios de unión 2 y 4, próximos a los sitios 1 y 3 donde se une p4 de forma imprescindible para ejercer su función, permitiría un aumento de la concentración local de la proteína. Esto aseguraría la disponibilidad de p4, favoreciendo la unión a sus sitios y evitando la dilución de la proteína sintetizada tras la transcripción temprana.

Los ensayos de cooperatividad en la unión de p4 a sus sitios apoyan esta propuesta. Los resultados indican que, aunque la cooperatividad es solo de 4-7 veces, la unión de p4 es mejor cuando están presentes dos sitios que cuando solamente hay uno; este hecho está de acuerdo con la función propuesta, puesto que se requiere un cierto aumento de la afinidad para asegurar la disponibilidad de la proteína en la secuencia a la que regula. Por otra parte, es importante que estos sitios de almacenamiento de p4 no sean sitios fuertes de unión, porque p4 debe poder ser fácilmente desplazada del sitio 2, para la correcta formación del complejo p4-p6, y del sitio 4, para la unión de la RNAP en el promotor A3; además, debe poder ser recolocada rápidamente de los sitios 2 y 4 a los sitios 1 y 3 cuando sea necesario.

Hemos visto que la interacción de las proteínas p4N y p6N de Nf con el DNA parece más fuerte que la de las proteínas de  $\phi 29$ ; además, la formación de los complejos p4N-p6N en el DNA de  $\phi 29$  se forman y funcionan, mientras que en el DNA de Nf los complejos con proteínas de  $\phi 29$  se forman difícilmente y no son, en ningún caso, funcionales. Estos hechos podrían justificar la ausencia de los sitios 2 y 4 en el genoma de Nf, ya que al ser mayor la afinidad de p4 por el DNA no es tan necesaria la presencia de sitios adicionales para el reclutamiento de p4 y ha podido perderlos más fácilmente.

Dado que la regulación de la transcripción está mediada por el complejo formado por las proteínas p4 y p6 y que la orientación de los monómeros de p4 en los sitios 1 y 3 de  $\phi 29$  es BA-AB, los monómeros de unión más débil al DNA (monómeros A), están orientados hacia la secuencia donde se une p6. Pensamos que en la formación del complejo regulador, los dímeros de p4 interaccionarían primero a través de los monómeros B en los sitios 1 y 3 y posicionarían a p6 en la secuencia comprendida entre ambos. La proteína p6 estabilizaría, a su vez, la unión de p4 a través de la interacción con sus monómeros A (figura 70). Y esta sería la base de la sinergia en la formación del complejo p4-p6-DNA.

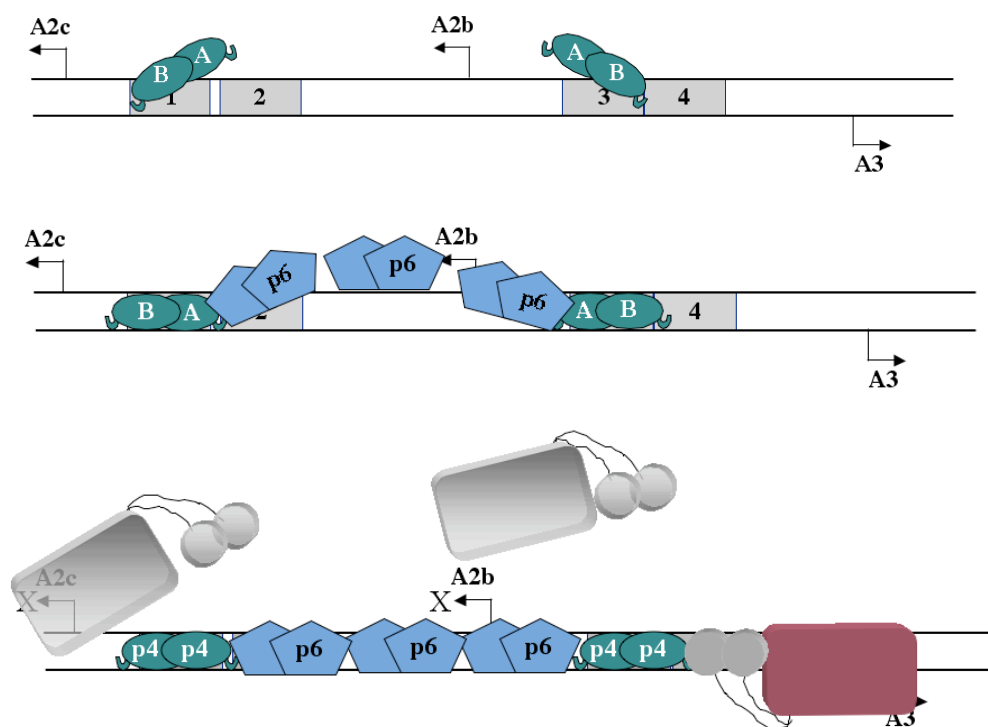


Figura 70: Esquema del modelo propuesto para la formación del complejo funcional p4-p6 que regula el paso de la transcripción temprana a la tardía en el fago  $\phi 29$ .

Cuando se estudió el mecanismo de regulación de la transcripción de Nf, a pesar de ser muy similar al de  $\phi 29$ , encontramos algunas diferencias importantes, entre las que podríamos destacar que la RNAP de *B. subtilis* es capaz de transcribir, aunque de forma débil, el promotor tardío A3 en ausencia de p4; por otra parte, la p4 de Nf presenta mayor afinidad por su sitio 1 que por su sitio 3 y además carece de los sitios 2 y 4 de unión, sugiriendo que Nf ha evolucionado para asegurarse una alta afinidad y una fuerte unión de la p4N a sus sitios.

La orientación de los monómeros de p4, en el caso de Nf, no es fácilmente predecible; en el sitio 3 el monómero B, de unión más fuerte, probablemente es el que está orientado hacia la secuencia donde se une la p6, al contrario de lo que ocurre en  $\phi 29$ , y en cuanto al sitio 1, cualquiera de los dos extremos podría ser el de mayor afinidad. Esto indicaría que la formación del complejo p4N-p6N es posiblemente distinta a la que hemos propuesto para la formación del complejo p4-p6 de  $\phi 29$ , lo cual concuerda con el hecho de que las proteínas de ambos sistemas no sean intercambiables.



#### **4.- GENERALIDAD Y DIVERSIDAD DEL MECANISMO DE REGULACIÓN DE LA TRANSCRIPCIÓN**

Los resultados obtenidos en el estudio de la regulación de la transcripción del fago Nf nos han permitido corroborar el hecho de que para que se produzca el paso de la transcripción temprana a la transcripción tardía, es necesaria la interacción funcional en el DNA de los promotores A2c, A2b y A3, de las proteínas p4 y p6, formando un complejo nucleoproteico que implica la unión de la proteína p4 a sus sitios 1 y 3 y la unión de p6 a lo largo de la secuencia de DNA comprendida entre ambos. Ninguna de las combinaciones de las proteínas de  $\phi 29$  con o sin proteínas de Nf, fue capaz de funcionar adecuadamente en el DNA de Nf, en las condiciones analizadas. El complejo p4-p6 en el DNA de Nf se formaba con proteínas de  $\phi 29$  pero era necesario emplear ocho veces más cantidad de proteína p6 y no era funcional. Pudimos observar también que las proteínas p4N y p6N se unen de forma sinérgica al DNA de Nf, del mismo modo que la p4 y la p6 lo hacen en el DNA de  $\phi 29$  (Camacho y Salas, 2001). El patrón de la digestión con DNasa I que se genera de los complejos de las proteínas de  $\phi 29$  y de Nf son muy similares; aun así la p4 y p6 de  $\phi 29$ , en el DNA de Nf no se unen sinérgicamente. Esto está indicando que en  $\phi 29$  la proteína p4 específicamente reconoce a la proteína p6 homóloga y viceversa, y que se requiere un posicionamiento muy preciso de las proteínas en el DNA. Este hecho está en línea con estudios de la replicación de  $\phi 29$  y de Nf, en los que se vio que la iniciación de la replicación solo se producía cuando todos los componentes del proceso eran del mismo fago (Bravo *et al.*, 1994; Freire *et al.*, 1996; González-Huici *et al.*, 2000).

La regulación de la transcripción es, por tanto, muy similar en ambos fagos, pero las proteínas implicadas no son intercambiables en los dos sistemas, lo que podría explicarse por las diferencias encontradas en las afinidades de la proteína p4, en la orientación de los monómeros, en la formación de complejo con la p6, etc.

Todos estos datos parecen indicar que estos dos fagos relacionados han evolucionado y se han adaptado para coordinar funcionalmente la regulación de la transcripción de los promotores de sus sistemas, utilizando sus propias proteínas virales. De modo que un mismo mecanismo se ha seleccionado evolutivamente como óptimo al menos en dos organismos diferentes, que a su vez se ha diferenciado lo suficiente para hacerse específico en cada uno de ellos.

# *CONCLUSIONES*

- 1.- El sitio mínimo de unión de la proteína p4 en el DNA contiene tres *huellas* consecutivas, determinadas por tratamiento con radicales hidroxilo, y coincide con una repetición invertida en la secuencia. Hay cuatro sitios de unión de p4 entre el promotor temprano A2c y el tardío A3.
- 2.- La proteína p4 no tiene la misma afinidad relativa por sus cuatro sitios de unión, presentando el doble de afinidad por el sitio 3 que por el sitio 1, y cinco veces más por el sitio 1 que por el 2. El sitio 4 es el de menor afinidad.
- 3.- Para unirse al DNA, la proteína p4 no requiere la existencia previa de una curvatura en la secuencia. La p4 induce una curvatura alrededor de 86° cuando se une al DNA que contiene una región de unión.
- 4.- La proteína p4 se une al DNA como dímero. Un sitio de unión de p4 se caracteriza por ser una secuencia de 27 pb, que posee dos guaninas, una en cada cadena, separadas por 25 nucleótidos; tiene tres regiones de pares de bases A/T que le aportan flexibilidad y que permiten que se genere la curvatura del sitio de unión.
- 5.- La proteína p4 establece únicamente una interacción *específica directa* entre un aminoácido y una base, mediada por la cadena lateral de las argininas en posición 6 de cada monómero con la base nitrogenada de las guaninas en las posiciones +/-13 del sitio de unión. La formación del complejo p4-DNA depende, además, del reconocimiento *específico indirecto* de las secuencias de A/T consecutivas.
- 6.- Los aminoácidos Arg6, Thr4, Tyr33 y Lys76 de p4 son críticos para la unión de la proteína al DNA. La His10 y la Lys36 son necesarias para la adecuada formación de los complejos funcionales implicados en la regulación de la transcripción. La Lys51 y la Arg54 no son esenciales para formación del complejo, pero contribuyen a su estabilidad. La Gln5 no está implicada en la unión de p4 al DNA.
- 7.- Los monómeros de p4 tienen distinta afinidad por las secuencias de los dos extremos del sitio de unión.
- 8.- La entidad funcional que regula el paso de la transcripción temprana a la transcripción tardía en el ciclo vital de  $\phi 29$  es un complejo p4-p6-RNAP, que implica la unión de la p4 en los sitios 1 y 3 y de p6 en la secuencia comprendida entre ambos. Los sitios 2 y 4 de unión de p4 no tienen un papel directo en la regulación de la transcripción del DNA de  $\phi 29$ .
- 9.- La regulación de la transcripción del bacteriófago Nf sigue un mecanismo muy parecido al de  $\phi 29$ , mediado por proteínas homólogas a p4 y a p6. Aunque existe gran similitud en las secuencias de las proteínas y de los promotores, el proceso de transcripción en Nf no es regulado por las proteínas de  $\phi 29$ , lo que hace a este sistema regulador, a su vez, específico de cada organismo.

# *BIBLIOGRAFÍA*

- Abril, A. M., Salas, M., Andreu, J. M., Hermoso, J. M. y Rivas, G. (1997). Phage  $\phi$ 29 protein p6 is in a monomer-dimer equilibrium that shifts to higher association states at the millimolar concentrations found in vivo. *Biochemistry* **36**, 11901-11908.
- Aiyar, S. E., Gourse, R. L. y Ross, W. (1998). Upstream A-tracts increase bacterial promoter activity through interactions with the RNA polymerase  $\alpha$  subunit. *Proc Natl Acad Sci U S A* **95**, 14652-14657.
- Aiyar, S. E., Juang, Y. L., Helmann, J. D. y deHaseth, P. L. (1994). Mutations in  $\sigma$  factor that affect the temperature dependence of transcription from a promoter, but not from a mismatch bubble in double-stranded DNA. *Biochemistry* **33**, 11501-11506.
- Allison, L. A., Moyle, M., Shales, M. y Ingles, C. J. (1985). Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell* **42**, 599-610.
- Badia, D., Camacho, A., Pérez-Lago, L., Escandón, C., Salas, M. y Coll, M. (2006). The structure of phage  $\phi$ 29 transcription regulator p4-DNA complex reveals an N-hook motif for DNA binding. *Mol. Cell* **22**, 73-81.
- Barker, M. M., Gaal, T., Josaitis, C. A. y Gourse, R. L. (2001). Mechanism of regulation of transcription initiation by ppGpp. I. Effects of ppGpp on transcription initiation in vivo and in vitro. *J Mol Biol* **305**, 673-688.
- Barnard, A., Wolfe, A. y Busby, S. (2004). Regulation at complex bacterial promoters: how bacteria use different promoter organizations to produce different regulatory outcomes. *Curr Opin Microbiol* **7**, 102-108.
- Barne, K. A., Bown, J. A., Busby, S. J. y Minchin, S. D. (1997). Region 2.5 of the *Escherichia coli* RNA polymerase  $\sigma$ 70 subunit is responsible for the recognition of the 'extended-10' motif at promoters. *EMBO J* **16**, 4034-4040.
- Barthelemy, I., Lázaro, J. M., Méndez, E., Mellado, R. P. y Salas, M. (1987). Purification in an active form of the phage  $\phi$ 29 protein p4 that controls the viral late transcription. *Nucleic Acids Res* **15**, 7781-7793.
- Barthelemy, I., Mellado, R. P. y Salas, M. (1988). Symmetrical transcription in bacteriophage  $\phi$ 29 DNA. *Biochimie* **70**, 605-609.
- Barthelemy, I., Mellado, R. P. y Salas, M. (1989). In vitro transcription of bacteriophage  $\phi$ 29 DNA: inhibition of early promoters by the viral replication protein p6. *J Virol* **63**, 460-462.
- Barthelemy, I. y Salas, M. (1989). Characterization of a new prokaryotic transcriptional activator and its DNA recognition site. *J Mol Biol* **208**, 225-232.
- Beatty, C. M., Browning, D. F., Busby, S. J. y Wolfe, A. J. (2003). Cyclic AMP receptor protein-dependent activation of the *Escherichia coli* acsP2 promoter by a synergistic class III mechanism. *J Bacteriol* **185**, 5148-5157.
- Belyaeva, T., Griffiths, L., Minchin, S., Cole, J. y Busby, S. (1993). The *Escherichia coli* cysG promoter belongs to the 'extended -10' class of bacterial promoters. *Biochem J* **296** (Pt 3), 851-857.

- Benoff, B., Yang, H., Lawson, C. L., Parkinson, G., Liu, J., Blatter, E., Ebright, Y. W., Berman, H. M. y Ebright, R. H. (2002). Structural basis of transcription activation: the CAP- $\alpha$ -CTD-DNA complex. *Science* **297**, 1562-1566.
- Blatter, E. E., Ross, W., Tang, H., Gourse, R. L. y Ebright, R. H. (1994). Domain organization of RNA polymerase  $\alpha$  subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. *Cell* **78**, 889-896.
- Borukhov, S. y Lee, J. (2005). RNA polymerase structure and function at lac operon. *C R Biol* **328**, 576-587.
- Borukhov, S. y Nudler, E. (2003). RNA polymerase holoenzyme: structure, function and biological implications. *Curr Opin Microbiol* **6**, 93-100.
- Bowers, C. W. y Dombroski, A. J. (1999). A mutation in region 1.1 of  $\sigma$ 70 affects promoter DNA binding by *Escherichia coli* RNA polymerase holoenzyme. *EMBO J* **18**, 709-716.
- Bravo, A., Hermoso, J. M. y Salas, M. (1994). In vivo functional relationships among terminal proteins of *Bacillus subtilis*  $\phi$ 29-related phages. *Gene* **148**, 107-112.
- Browning, D. F., Beatty, C. M., Wolfe, A. J., Cole, J. A. y Busby, S. J. (2002). Independent regulation of the divergent *Escherichia coli* nrfA and acsP1 promoters by a nucleoprotein assembly at a shared regulatory region. *Mol Microbiol* **43**, 687-701.
- Browning, D. F. y Busby, S. J. (2004). The regulation of bacterial transcription initiation. *Nat Rev Microbiol* **2**, 57-65.
- Burgess, R. R. (1969). Separation and characterization of the subunits of ribonucleic acid polymerase. *J Biol Chem* **244**, 6168-6176.
- Burgess, R. R., Travers, A. A., Dunn, J. J. y Bautz, E. K. (1969). Factor stimulating transcription by RNA polymerase. *Nature* **221**, 43-46.
- Busby, S. y Ebright, R. H. (1994). Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell* **79**, 743-746.
- Busby, S. y Ebright, R. H. (1997). Transcription activation at class II CAP-dependent promoters. *Mol Microbiol* **23**, 853-859.
- Busby, S. y Ebright, R. H. (1999). Transcription activation by catabolite activator protein (CAP). *J Mol Biol* **293**, 199-213.
- Calles, B., Salas, M. y Rojo, F. (2002). The  $\phi$ 29 transcriptional regulator contacts the nucleoid protein p6 to organize a repression complex. *EMBO J* **21**, 6185-6194.
- Camacho, A. y Salas, M. (1999). Effect of mutations in the "extended -10" motif of three *Bacillus subtilis*  $\sigma$ A-RNA polymerase-dependent promoters. *J Mol Biol* **286**, 683-693.
- Camacho, A. y Salas, M. (2000). Pleiotropic effect of protein P6 on the viral cycle of bacteriophage  $\phi$ 29. *J Bacteriol* **182**, 6927-6932.
- Camacho, A. y Salas, M. (2001). Mechanism for the switch of  $\phi$ 29 DNA early to late transcription by regulatory protein p4 and histone-like protein p6. *EMBO J* **20**, 6060-6070.

- Camacho, A. y Salas, M. (2004). Molecular interplay between RNA polymerase and two transcriptional regulators in promoter switch. *J Mol Biol* **336**, 357-368.
- Camarero, J. A., Shekhtman, A., Campbell, E. A., Chlenov, M., Gruber, T. M., Bryant, D. A., Darst, S. A., Cowburn, D. y Muir, T. W. (2002). Autoregulation of a bacterial sigma factor explored by using segmental isotopic labeling and NMR. *Proc Natl Acad Sci U S A* **99**, 8536-8541.
- Campbell, E. A., Muzzin, O., Chlenov, M., Sun, J. L., Olson, C. A., Weinman, O., Trester-Zedlitz, M. L. y Darst, S. A. (2002). Structure of the bacterial RNA polymerase promoter specificity  $\sigma$  subunit. *Mol Cell* **9**, 527-539.
- Campbell, E. A., Tupy, J. L., Gruber, T. M., Wang, S., Sharp, M. M., Gross, C. A. y Darst, S. A. (2003). Crystal structure of *Escherichia coli*  $\sigma$ E with the cytoplasmic domain of its anti-sigma RseA. *Mol Cell* **11**, 1067-1078.
- Carrascosa, J. L., Camacho, A., Moreno, F., Jiménez, F., Mellado, R. P., Viñuela, E. y Salas, M. (1976). *Bacillus subtilis* phage  $\phi$ 29. Characterization of gene products and functions. *Eur J Biochem* **66**, 229-241.
- Chatterji, D. y Ojha, A. K. (2001). Revisiting the stringent response, ppGpp and starvation signaling. *Curr Opin Microbiol* **4**, 160-165.
- Chen, H., Tang, H. y Ebright, R. H. (2003). Functional interaction between RNA polymerase  $\alpha$  subunit C-terminal domain and  $\sigma$ 70 in UP-element- and activator-dependent transcription. *Mol Cell* **11**, 1621-1633.
- Crothers, D. M. y Drak, J. (1992). Global features of DNA structure by comparative gel electrophoresis. *Methods Enzymol* **212**, 46-71.
- Dame, R. T., Wyman, C. y Goosen, N. (2001). Structural basis for preferential binding of H-NS to curved DNA. *Biochimie* **83**, 231-234.
- Darst, S. A. (2001). Bacterial RNA polymerase. *Curr Opin Struct Biol* **11**, 155-162.
- Darst, S. A., Opalka, N., Chacon, P., Polyakov, A., Richter, C., Zhang, G. y Wriggers, W. (2002). Conformational flexibility of bacterial RNA polymerase. *Proc Natl Acad Sci U S A* **99**, 4296-4301.
- Demple, B. (1996). Redox signaling and gene control in the *Escherichia coli* soxRS oxidative stress regulon--a review. *Gene* **179**, 53-57.
- Dickerson, R. E. y Chiu, T. K. (1997). Helix bending as a factor in protein/DNA recognition. *Biopolymers* **44**, 361-403.
- Dieckman, S. (1989). The migration anomaly of DNA fragments in polyacrylamide gels allows the detection of small sequence-specific DNA structure variations. *Electrophoresis* **10**, 354-359.
- Dorman, C. J. y Deighan, P. (2003). Regulation of gene expression by histone-like proteins in bacteria. *Curr Opin Genet Dev* **13**, 179-184.
- Dove, S. L., Darst, S. A. y Hochschild, A. (2003). Region 4 of sigma as a target for transcription regulation. *Mol Microbiol* **48**, 863-874.

- Ebright, R. H. (1993). Transcription activation at Class I CAP-dependent promoters. *Mol Microbiol* **8**, 797-802.
- Ebright, R. H. (2000). RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J Mol Biol* **304**, 687-698.
- Ebright, R. H. y Busby, S. (1995). The *Escherichia coli* RNA polymerase alpha subunit: structure and function. *Curr Opin Genet Dev* **5**, 197-203.
- Elías-Arnanz, M. y Salas, M. (1999). Functional interactions between a phage histone-like protein and a transcriptional factor in regulation of  $\phi$ 29 early-late transcriptional switch. *Genes Dev* **13**, 2502-2513.
- Ellinger, T., Behnke, D., Bujard, H. y Gralla, J. D. (1994). Stalling of *Escherichia coli* RNA polymerase in the +6 to +12 region in vivo is associated with tight binding to consensus promoter elements. *J Mol Biol* **239**, 455-465.
- Escarmís, C., Guirao, D. y Salas, M. (1989). Replication of recombinant  $\phi$ 29 DNA molecules in *Bacillus subtilis* protoplasts. *Virology* **169**, 152-160.
- Falconi, M., Prosseda, G., Giangrossi, M., Beghetto, E. y Colonna, B. (2001). Involvement of FIS in the H-NS-mediated regulation of virF gene of *Shigella* and enteroinvasive *Escherichia coli*. *Mol Microbiol* **42**, 439-452.
- Freire, R., Salas, M. y Hermoso, J. M. (1994). A new protein domain for binding to DNA through the minor groove. *EMBO J* **13**, 4353-4360.
- Freire, R., Serrano, M., Salas, M. y Hermoso, J. M. (1996). Activation of replication origins in  $\phi$ 29-related phages requires the recognition of initiation proteins to specific nucleoprotein complexes. *J Biol Chem* **271**, 31000-31007.
- González-Huici, V., Lázaro, J. M., Salas, M. y Hermoso, J. M. (2000). Specific recognition of parental terminal protein by DNA polymerase for initiation of protein-primed DNA replication. *J Biol Chem* **275**, 14678-14683.
- González-Huici, V., Salas, M. y Hermoso, J. M. (2006). Requirements for *Bacillus subtilis* bacteriophage  $\phi$ 29 DNA ejection. *Gene* **7**, 19-25.
- Gourse, R. L., Ross, W. y Gaal, T. (2000). UPs and downs in bacterial transcription initiation: the role of the  $\alpha$  subunit of RNA polymerase in promoter recognition. *Mol Microbiol* **37**, 687-695.
- Grainger, D. C., Webster, C. L., Belyaeva, T. A., Hyde, E. I. y Busby, S. J. (2004). Transcription activation at the *Escherichia coli* melAB promoter: interactions of MelR with its DNA target site and with domain 4 of the RNA polymerase sigma subunit. *Mol Microbiol* **51**, 1297-1309.
- Gruber, T. M. y Gross, C. A. (2003). Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu Rev Microbiol* **57**, 441-466.
- Gruber, T. M., Markov, D., Sharp, M. M., Young, B. A., Lu, C. Z., Zhong, H. J., Artsimovitch, I., Geszvain, K. M., Arthur, T. M., Burgess, R. R., et al. (2001). Binding of the initiation factor sigma(70) to core RNA polymerase is a multistep process. *Mol Cell* **8**, 21-31.



- Gutiérrez, C., Freire, R., Salas, M. y Hermoso, J. M. (1994). Assembly of phage  $\phi$ 29 genome with viral protein p6 into a compact complex. *EMBO J* **13**, 269-276.
- Helmann, J. D. (1999). Anti-sigma factors. *Curr Opin Microbiol* **2**, 135-141.
- Hughes, K. T. y Mathee, K. (1998). The anti-sigma factors. *Annu Rev Microbiol* **52**, 231-286.
- Ishihama, A. (2000). Functional modulation of *Escherichia coli* RNA polymerase. *Annu Rev Microbiol* **54**, 499-518.
- Jeon, Y. H., Yamazaki, T., Otomo, T., Ishihama, A. y Kyogoku, Y. (1997). Flexible linker in the RNA polymerase alpha subunit facilitates the independent motion of the C-terminal activator contact domain. *J Mol Biol* **267**, 953-962.
- Juang, Y. L. y Helmann, J. D. (1994a). The delta subunit of *Bacillus subtilis* RNA polymerase. An allosteric effector of the initiation and core-recycling phases of transcription. *J Mol Biol* **239**, 1-14.
- Juang, Y. L. y Helmann, J. D. (1994b). A promoter melting region in the primary sigma factor of *Bacillus subtilis*. Identification of functionally important aromatic amino acids. *J Mol Biol* **235**, 1470-1488.
- Kahn, J. D. (1999). Methods for analyzing DNA bending. *Methods Mol Biol* **94**, 109-123.
- Kar, S. y Adhya, S. (2001). Recruitment of HU by piggyback: a special role of GalR in repressosome assembly. *Genes Dev* **15**, 2273-2281.
- Katahira, M., Sugeta, H. y Kyogoku, Y. (1990). A new model for the bending of DNAs containing the oligo(dA) tracts based on NMR observations. *Nucleic Acids Res* **18**, 613-618.
- Keilty, S. y Rosenberg, M. (1987). Constitutive function of a positively regulated promoter reveals new sequences essential for activity. *J Biol Chem* **262**, 6389-6395.
- Kim, J., Zwieb, C., Wu, C. y Adhya, S. (1989). Bending of DNA by gene-regulatory proteins: construction and use of a DNA bending vector. *Gene* **85**, 15-23.
- Komissarova, N. y Kashlev, M. (1997). Transcriptional arrest: *Escherichia coli* RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. *Proc Natl Acad Sci U S A* **94**, 1755-1760.
- Kopka, M. L., Fratini, A. V., Drew, H. R. y Dickerson, R. E. (1983). Ordered water structure around a B-DNA dodecamer. A quantitative study. *J Mol Biol* **163**, 129-146.
- Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A. y Darst, S. A. (2000). A structural model of transcription elongation. *Science* **289**, 619-625.
- Kumar, A., Malloch, R. A., Fujita, N., Smillie, D. A., Ishihama, A. y Hayward, R. S. (1993). The minus 35-recognition region of *Escherichia coli*  $\sigma$ 70 is inessential for initiation of transcription at an "extended minus 10" promoter. *J Mol Biol* **232**, 406-418.
- Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.
- Lisser, S. y Margalit, H. (1993). Compilation of *E. coli* mRNA promoter sequences. *Nucleic Acids Res* **21**, 1507-1516.

- Liu, M., Tolstorukov, M., Zhurkin, V., Garges, S. y Adhya, S. (2004). A mutant spacer sequence between -35 and -10 elements makes the Plac promoter hyperactive and cAMP receptor protein-independent. *Proc Natl Acad Sci U S A* **101**, 6911-6916.
- Lonetto, M., Gribskov, M. y Gross, C. A. (1992). The  $\sigma$  70 family: sequence conservation and evolutionary relationships. *J Bacteriol* **174**, 3843-3849.
- López de Saro, F. J., Yoshikawa, N. y Helmann, J. D. (1999). Expression, abundance, and RNA polymerase binding properties of the delta factor of *Bacillus subtilis*. *J Biol Chem* **274**, 15953-15958.
- MacDonald, D., Herbert, K., Zhang, X., Pologruto, T. y Lu, P. (2001). Solution structure of an A-tract DNA bend. *J Mol Biol* **306**, 1081-1098.
- Mathew, R. y Chatterji, D. (2006). The evolving story of the omega subunit of bacterial RNA polymerase. *Trends Microbiol.* **14**, 450-455.
- Maxam, A. M. y Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol* **65**, 499-560.
- McClure, W. R. (1985). Mechanism and control of transcription initiation in prokaryotes. *Annu Rev Biochem* **54**, 171-204.
- Mekler, V., Kortkhonjia, E., Mukhopadhyay, J., Knight, J., Revyakin, A., Kapanidis, A. N., Niu, W., Ebright, Y. W., Levy, R. y Ebright, R. H. (2002). Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell* **108**, 599-614.
- Mencía, M., Monsalve, M., Rojo, F. y Salas, M. (1996). Transcription activation by phage  $\phi$ 29 protein p4 is mediated by interaction with the  $\alpha$  subunit of *Bacillus subtilis* RNA polymerase. *Proc Natl Acad Sci U S A* **93**, 6616-6620.
- Mencía, M., Salas, M. y Rojo, F. (1993). Residues of the *Bacillus subtilis* phage  $\phi$ 29 transcriptional activator required both to interact with RNA polymerase and to activate transcription. *J Mol Biol* **233**, 695-704.
- Mendieta, J., Pérez-Lago, L., Salas, M. (2007). DNA sequence- specific recognition by transcriptional regulator through indirect readout of A-tracts. Enviado para publicar.
- Minakhin, L., Bhagat, S., Brunning, A., Campbell, E. A., Darst, S. A., Ebright, R. H. y Severinov, K. (2001). Bacterial RNA polymerase subunit omega and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. *Proc Natl Acad Sci U S A* **98**, 892-897.
- Mizukami, Y., Sekiya, T. y Hirokawa, H. (1986). Nucleotide sequence of gene F of *Bacillus* phage Nf. *Gene* **42**, 231-235.
- Monsalve, M., Calles, B., Mencía, M., Rojo, F. y Salas, M. (1998). Binding of phage  $\phi$ 29 protein p4 to the early A2c promoter: recruitment of a repressor by the RNA polymerase. *J Mol Biol* **283**, 559-569.
- Monsalve, M., Mencía, M., Rojo, F. y Salas, M. (1995). Transcription regulation in *Bacillus subtilis* phage  $\phi$ 29: expression of the viral promoters throughout the infection cycle. *Virology* **207**, 23-31.

- Monsalve, M., Mencía, M., Salas, M. y Rojo, F. (1996). Protein p4 represses phage ø29 A2c promoter by interacting with the alpha subunit of *Bacillus subtilis* RNA polymerase. *Proc Natl Acad Sci U S A* **93**, 8913-8918.
- Mooney, R. A., Darst, S. A. y Landick, R. (2005). Sigma and RNA polymerase: an on-again, off-again relationship? *Mol Cell* **20**, 335-345.
- Moreno, F., Camacho A., Viñuela E. y Salas, M. (1974). Suppressor-sensitive mutants and genetic map of *Bacillus subtilis* bacteriophage ø29. *Virology* **62**, 1-16.
- Muller-Hill, B. (1998). Some repressors of bacterial transcription. *Curr Opin Microbiol* **1**, 145-151.
- Murakami, K., Fujita, N. y Ishihama, A. (1996). Transcription factor recognition surface on the RNA polymerase alpha subunit is involved in contact with the DNA enhancer element. *EMBO J* **15**, 4358-4367.
- Murakami, K. S. y Darst, S. A. (2003). Bacterial RNA polymerases: the whole story. *Curr Opin Struct Biol* **13**, 31-39.
- Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O. y Darst, S. A. (2002). Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* **296**, 1285-1290.
- Negishi, T., Fujita, N. y Ishihama, A. (1995). Structural map of the alpha subunit of *Escherichia coli* RNA polymerase: structural domains identified by proteolytic cleavage. *J Mol Biol* **248**, 723-728.
- Nuez, B. y Salas, M. (1993). Bacteriophage Nf DNA region controlling late transcription: structural and functional homology with bacteriophage ø29. *Nucleic Acids Res* **21**, 2861-2865.
- Olson, W. K., Gorin, A. A., Lu, X. J., Hock, L. M. y Zhurkin, V. B. (1998). DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc Natl Acad Sci U S A* **95**, 11163-11168.
- Owens, J. T., Miyake, R., Murakami, K., Chmura, A. J., Fujita, N., Ishihama, A. y Meares, C. F. (1998). Mapping the sigma70 subunit contact sites on *Escherichia coli* RNA polymerase with a sigma70-conjugated chemical protease. *Proc Natl Acad Sci U S A* **95**, 6021-6026.
- Pastrana, R., Lázaro, J. M., Blanco, L., García, J. A., Méndez, E. y Salas, M. (1985). Overproduction and purification of protein p6 of *Bacillus subtilis* phage ø29: role in the initiation of DNA replication. *Nucleic Acids Res* **13**, 3083-3100.
- Pecenкова, T. y Paces, V. (1999). Molecular phylogeny of ø29-like phages and their evolutionary relatedness to other protein-primed replicating phages and other phages hosted by gram-positive bacteria. *J Mol Evol* **48**, 197-208.
- Pérez-Martín, J. y Espinosa, M. (1994). Correlation between DNA bending and transcriptional activation at a plasmid promoter. *J Mol Biol* **241**, 7-17.
- Pérez-Rueda, E. y Collado-Vides, J. (2000). The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. *Nucleic Acids Res* **28**, 1838-1847.

- Richet, E., Vidal-Ingigliardi, D. y Raibaud, O. (1991). A new mechanism for coactivation of transcription initiation: repositioning of an activator triggered by the binding of a second activator. *Cell* **66**, 1185-1195.
- Rojo, F., Zaballos, A. y Salas, M. (1990). Bend induced by the phage  $\phi$ 29 transcriptional activator in the viral late promoter is required for activation. *J Mol Biol* **211**, 713-725.
- Ross, W., Ernst, A. y Gourse, R. L. (2001). Fine structure of *E. coli* RNA polymerase-promoter interactions: alpha subunit binding to the UP element minor groove. *Genes Dev* **15**, 491-506.
- Ross, W., Gosink, K. K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K. y Gourse, R. L. (1993). A third recognition element in bacterial promoters: DNA binding by the  $\alpha$  subunit of RNA polymerase. *Science* **262**, 1407-1413.
- Ross, W., Schneider, D. A., Paul, B. J., Mertens, A. y Gourse, R. L. (2003). An intersubunit contact stimulating transcription initiation by *E coli* RNA polymerase: interaction of the  $\alpha$  C-terminal domain and  $\sigma$  region 4. *Genes Dev* **17**, 1293-1307.
- Salgado, H., Santos-Zavaleta, A., Gama-Castro, S., Millán-Zarate, D., Díaz-Peredo, E., Sánchez-Solano, F., Pérez-Rueda, E., Bonavides-Martínez, C. y Collado-Vides, J. (2001). RegulonDB (version 3.2): transcriptional regulation and operon organization in *Escherichia coli* K-12. *Nucleic Acids Res* **29**, 72-74.
- Sambrook, J., Fritsch, E. F. y Maniatis, T. (1989). Molecular Cloning: a Laboratory Manual 2<sup>a</sup> edit. (Harbor, C. S., ed.). 3 vols. Cold Spring Harbor Laboratory, N. Y.
- Sanderson, A., Mitchell, J. E., Minchin, S. D. y Busby, S. J. (2003). Substitutions in the *Escherichia coli* RNA polymerase  $\sigma$ 70 factor that affect recognition of extended -10 elements at promoters. *FEBS Lett* **544**, 199-205.
- Savery, N. J., Lloyd, G. S., Busby, S. J., Thomas, M. S., Ebright, R. H. y Gourse, R. L. (2002). Determinants of the C-terminal domain of the *Escherichia coli* RNA polymerase alpha subunit important for transcription at class I cyclic AMP receptor protein-dependent promoters. *J Bacteriol* **184**, 2273-2280.
- Savery, N. J., Lloyd, G. S., Kainz, M., Gaal, T., Ross, W., Ebright, R. H., Gourse, R. L. y Busby, S. J. (1998). Transcription activation at Class II CRP-dependent promoters: identification of determinants in the C-terminal domain of the RNA polymerase  $\alpha$  subunit. *EMBO J* **17**, 3439-3447.
- Savery, N. J., Rhodius, V. A., Wing, H. J. y Busby, S. J. (1995). Transcription activation at *Escherichia coli* promoters dependent on the cyclic AMP receptor protein: effects of binding sequences for the RNA polymerase  $\alpha$ -subunit. *Biochem J* **309** (Pt 1), 77-83.
- Schneider, R., Travers, A., Kutateladze, T. y Muskhelishvili, G. (1999). A DNA architectural protein couples cellular physiology and DNA topology in *Escherichia coli*. *Mol Microbiol* **34**, 953-964.
- Schroder, O. y Wagner, R. (2000). The bacterial DNA-binding protein H-NS represses ribosomal RNA transcription by trapping RNA polymerase in the initiation complex. *J Mol Biol* **298**, 737-748.

- Semsey, S., Geanakopoulou, M., Lewis, D. E. y Adhya, S. (2002). Operator-bound GalR dimers close DNA loops by direct interaction: tetramerization and inducer binding. *EMBO J* **21**, 4349-4356.
- Serrano, M., Barthelemy, I. y Salas, M. (1991). Transcription activation at a distance by phage  $\phi 29$  protein p4. Effect of bent and non-bent intervening DNA sequences. *J Mol Biol* **219**, 403-414.
- Serrano, M., Gutiérrez, C., Salas, M. y Hermoso, J. M. (1993). Superhelical path of the DNA in the nucleoprotein complex that activates the initiation of phage  $\phi 29$  DNA replication. *J Mol Biol* **230**, 248-259.
- Serrano, M., Salas, M. y Hermoso, J. M. (1990). A novel nucleoprotein complex at a replication origin. *Science* **248**, 1012-1016.
- Severinov, K. (2000). RNA polymerase structure-function: insights into points of transcriptional regulation. *Curr Opin Microbiol* **3**, 118-125.
- Shanil, P.H., Berkmen, M.B., Ross, W., Gaal, T. Ward, C. y Gourse, R. L. (2006). rRNA promoter regulation by nonoptimal binding of  $\sigma$  region 1.2: an additional recognition element for RNA polymerase. *Cell* **125**, 1069-1082.
- Sharp, M. M., Chan, C. L., Lu, C. Z., Marr, M. T., Nechaev, S., Merritt, E. W., Severinov, K., Roberts, J. W. y Gross, C. A. (1999). The interface of sigma with *core* RNA polymerase is extensive, conserved, and functionally specialized. *Genes Dev* **13**, 3015-3026.
- Sheridan, S. D., Benham, C. J. y Hatfield, G. W. (1998). Activation of gene expression by a novel DNA structural transmission mechanism that requires supercoiling-induced DNA duplex destabilization in an upstream activating sequence. *J Biol Chem* **273**, 21298-21308.
- Sheridan, S. D., Opel, M. L. y Hatfield, G. W. (2001). Activation and repression of transcription initiation by a distant DNA structural transition. *Mol Microbiol* **40**, 684-690.
- Shin, M., Kang, S., Hyun, S. J., Fujita, N., Ishihama, A., Valentin-Hansen, P. y Choy, H. E. (2001). Repression of deoP2 in *Escherichia coli* by CytR: conversion of a transcription activator into a repressor. *Embo J* **20**, 5392-5399.
- Sogo, J. M., Inciarte, M. R., Corral, J., Viñuela, E. y Salas, M. (1979). RNA polymerase binding sites and transcription map of the DNA of *Bacillus subtilis* phage  $\phi 29$ . *J Mol Biol* **127**, 411-436.
- Sorenson, M. K., Ray, S. S. y Darst, S. A. (2004). Crystal structure of the flagellar sigma/anti-sigma complex sigma(28)/FlgM reveals an intact sigma factor in an inactive conformation. *Mol Cell* **14**, 127-138.
- Stock, A. M., Robinson, V. L. y Goudreau, P. N. (2000). Two-component signal transduction. *Annu Rev Biochem* **69**, 183-215.
- Studier, F. W. y Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**, 113-130.
- Sweetser, D., Nonet, M. y Young, R. A. (1987). Prokaryotic and eukaryotic RNA polymerases have homologous *core* subunits. *Proc Natl Acad Sci U S A* **84**, 1192-1196.

- Thompson, J. F. y Landy, A. (1988). Empirical estimation of protein-induced DNA bending angles: applications to lambda site-specific recombination complexes. *Nucleic Acids Res* **16**, 9687-9705.
- Tomsic, M., Tsujikawa, L., Panaghie, G., Wang, Y., Azok, J. y deHaseth, P. L. (2001). Different roles for basic and aromatic amino acids in conserved region 2 of *Escherichia coli* sigma(70) in the nucleation and maintenance of the single-stranded DNA bubble in open RNA polymerase-promoter complexes. *J Biol Chem* **276**, 31891-31896.
- Tsujikawa, L., Tsodikov, O. V. y deHaseth, P. L. (2002). Interaction of RNA polymerase with forked DNA: evidence for two kinetically significant intermediates on the pathway to the final complex. *Proc Natl Acad Sci U S A* **99**, 3493-3498.
- Valentin-Hansen, P., Sogaard-Andersen, L. y Pedersen, H. (1996). A flexible partnership: the CytR anti-activator and the cAMP-CRP activator protein, comrades in transcription control. *Mol Microbiol* **20**, 461-466.
- Vassilyev, D. G., Sekine, S., Laptenko, O., Lee, J., Vassilyeva, M. N., Borukhov, S. y Yokoyama, S. (2002). Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* **417**, 712-719.
- Vrentas, C. E., Gaal, T., Ross, W., Ebright, R. H. y Gourse, R. L. (2005). Response of RNA polymerase to ppGpp: requirement for the omega subunit and relief of this requirement by DksA. *Genes Dev* **19**, 2378-2387.
- Vuthoori, S., Bowers, C. W., McCracken, A., Dombroski, A. J. y Hinton, D. M. (2001). Domain 1.1 of the sigma(70) subunit of *Escherichia coli* RNA polymerase modulates the formation of stable polymerase/promoter complexes. *J Mol Biol* **309**, 561-572.
- Wade, J. T., Belyaeva, T. A., Hyde, E. I. y Busby, S. J. (2001). A simple mechanism for co-dependence on two activators at an *Escherichia coli* promoter. *EMBO J* **20**, 7160-7167.
- Walker, K. A., Atkins, C. L. y Osuna, R. (1999). Functional determinants of the *Escherichia coli* fis promoter: roles of -35, -10, and transcription initiation regions in the response to stringent control and growth phase-dependent regulation. *J Bacteriol* **181**, 1269-1280.
- Westover, K. D., Bushnell, D. A. y Kornberg, R. D. (2004). Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* **119**, 481-489.
- Wilson, C. y Dombroski, A. J. (1997). Region 1 of sigma70 is required for efficient isomerization and initiation of transcription by *Escherichia coli* RNA polymerase. *J Mol Biol* **267**, 60-74.
- Wosten, M. M. (1998). Eubacterial sigma-factors. *FEMS Microbiol Rev* **22**, 127-150.
- Yoshikawa, H., K. J. Garvey, et al. (1985). Nucleotide sequence analysis of DNA replication origins of the small *Bacillus* bacteriophages: evolutionary relationships. *Gene* **37**(1-3): 125-30.
- Yoshikawa, H., Elder, J.H., Ito, J (1986). Comparative studies on the small *Bacillus subtilis* bacteriophages. *J Gen Appl Microbiol* **33**, 39-49.
- Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K. y Darst, S. A. (1999). Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* **98**, 811-824.

Zhang, G. y Darst, S. A. (1998). Structure of the *Escherichia coli* RNA polymerase  $\alpha$  subunit amino-terminal domain. *Science* **281**, 262-266.